

OS Escherichia coli O157:H7.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OC NCBI_Taxid=562, 83334;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12.
 RX MEDLINE-92114808; PubMed-1840644;
 RA "Sequencing, mutational analysis, and transcriptional regulation of
 RT the Escherichia coli htrb gene."
 RL Mol. Microbiol. 5:2285-2292(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / M61655;
 RX MEDLINE-97426617; PubMed-9278503;
 RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12."
 RL Science 277:1453-1474(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12.
 RX MEDLINE-97061202; PubMed-8905232;
 RA Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,
 RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kani K., Kashimoto K.,
 RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
 RA Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
 RA Sempel G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
 RA Yano M., Horinouchi T.;
 RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
 corresponding to the 12.7-28.0 min region on the linkage map."
 RL DNA Res. 3:137-155(1996).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE-21074935; PubMed-11206551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Postal G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grobbeck E.J., Davis N.W., Lam A., Dimalanta E.T., Potomousis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blatner F.R.;
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7."
 RL Nature 409:529-533(2001).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN-O157:H7 / RIMD 0509952;
 RX MEDLINE-21156231; PubMed-11258796;
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsuo E., Nakayama K., Murata T., Tanaka M., Tohe T.,
 RA Tida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
 RA Kudara S., Shiba T., Hattori M., Shinagawa H.;
 RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
 O157:H7 and genomic comparison with a laboratory strain K-12."
 RL DNA Res. 8:11-22(2001).
 RN [6]
 RP CHARACTERIZATION
 RC MEDLINE-91100364; PubMed-1846149;
 RA Karow M., Fayet O., Cegliska A., Zieglerhofer T., Georgopoulos C.;
 RT "Isolation and characterization of the Escherichia coli htrb gene,
 RT whose product is essential for bacterial viability above 33 degrees C
 RT in rich media."
 RL J. Bacteriol. 173:741-750(1991).
 RN [7]
 RP FUNCTION, AND CHARACTERIZATION
 RC MEDLINE-90368739; PubMed-2203778;
 RA Brozek K.A., Raelz C.R.H.;
 RT "Biosynthesis of lipid A in Escherichia coli. Acyl carrier protein-
 RT dependent incorporation of laurate and myristate."
 RL J. Biol. Chem. 265:15410-15417(1990).

CC -1- FUNCTION: ACYLATES THE INTERMEDIATE (KDO)2-LIPID IVA TO FORM
 CC (KDO)2-(LAUROYL)-LIPID IVA. HAS 10 FOLD SELECTIVITY FOR LAUROYL-
 CC ACP OVER MYRISTOYL-ACP.
 CC -1- PATHWAY: LIPOPOLYSACCHARIDE CORE BIOSYNTHESIS.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
 CC (potential).
 CC -1- INDUCTION: IS EXPRESSED AT ALL TEMPERATURES, BUT ACCUMULATION OF
 CC HTRB TRANSCRIPTS SLIGHTLY DECLINE WITH RAISING TEMPERATURE. THUS,
 CC ITS EXPRESSION IS NOT INDUCED BY HEAT SHOCK.
 CC -1- SIMILARITY: BELONGS TO THE HTRB/MSRB FAMILY.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X61000; CAA43317.1; -
 CC EMBL; AE000206; AAC74138.1; -
 CC EMBL; D90742; BAA35852.1; -
 CC EMBL; D90743; BAA35863.1; -
 CC EMBL; AE005316; AAG5800.1; -
 CC EMBL; AP002555; BAB34855.1; -
 CC PIR; S16888; S16888.
 CC ECGene; EG10464; htrb.
 CC Lipopolysaccharide biosynthesis; Transferrase; Acyltransferase;
 CC Transmembrane; Inner membrane; Heat shock; Complete proteome.
 CC TRANSMEM 17 37 POTENTIAL.
 CC TRANSMEM 209 229 POTENTIAL.
 CC SEQUENCE 306 AA; 35407 MW; AEA70A5BE10653B5 CRC64;
 SQ
 alignment_scores: Quality: 215.00 Length: 274
 Ratio: 1.303 Gaps: 9
 Percent Similarity: 60.219 Percent Identity: 29.197
 alignment_block:
 US-09-303-518D-569 x HTRB_ECOLI ..
 Align seg 1/1 to: HTRB_ECOLI from: 1 to: 306
 88 CTGCCCGCTTCTGCTGTCACACGCTGGGAACCGGCTGGACATCTGCG 137
 ||||| :||| ||||| ||||| |||||
 32 LeuprotyrProValIletryrArgleuglycysgilyleuyllysleuAl 48
 138 GTTTTACCTTTTAAAGAGACCGCGCGCATCTC...GCCATATATGC 184
 ||||| :||| ||||| ||||| |||||
 48 aleuAlgrpHemelllyArq...ArgAlalyslleValhlsArqAsnleug 64
 185 GT.....CAGCGACGATGATATCCGACCCCAAAACGGTCAAA 222
 :||| :||| :||| :||| :||| :|||
 64 luleucyspHeprGluMetserGluGluIuArqArgysMetValVal 80
 223 GCCGTTTTCGGAACCGCAAAAGCGGTTTGACACTTGCCCCCGCTT 272
 ||| :||| :||| :||| :||| :|||
 81 LysAsnphelGlservalGlyMetGlyLeuMetGluThrGlyMetLarr 97
 273 TTTTCAGAAACCGGAGACATAGAACATGTTCAAAAGCGGTACACGCT 322
 :||| :||| :||| :||| :||| :|||
 97 prHe...TriProAspargrAlleAlaArgtrPrThGluValIleGlyw 113
 323 GGGACATGTGACAGCGCTTTGGACAAACAGAGGCGCTATTCATC 372
 ||||| :||| :||| :||| :||| :|||
 113 etGluHsIleArqspValGlnAlaGlnLysArqGlylleLeuLeuVal 129
 373 ACGCGGCAATCGCGACGCTAGATTGGCGGACGCTACATCAGCAGCA 422
 ||||| :||| :||| :||| :||| :|||
 130 GlylleHsPheleuThrLeuGluLeuGlyAlaArgGlnphelGlyMetG 146
 423 GCTTCGCTCCCGTGCACCGCATGTACAAACCGCGCAAAATCAAGACGA 472

```

146 ngluProlyl.....ileglyValtyrArgProAsnAspAsnProLeuI 161
473 TAGACAAATATCATGACGGGAGGCGGTTGGCGGCAAAAGAAACCGCG 522
161 leAspTrpLeuGlnThrTrpPolylArgLeuArg.....SerAspLysSer 175
523 CCTACCAACATACAAAGGGTCAACAAATCATCAAGCCCTGCGTGGG 572
176 MetLeuAspArgLysAspLeuLysGlyMetIleLysAlaLeuLysGly 192
573 CGAACCAACCATCGCTGCGCCGACCGACGCTCCCTCCCTCAAGAGCG 622
192 ygluValAlaTrpTyrAlaProAspHisAspTyrGlyProArg..... 206
623 GGAAGCGGTATGGTGGTATTTCTTCGCG...AAACGTGCGCTATACCAT 669
207 ..SerSerValPheValProLeuPheAlaValGluGlnAlaThrThr 222
670 ACGTGGCGGCAAAATTTGGCACACGTCAAGCGCTGAAGACCTGTTT 719
223 ThrGlyThrTrpMetLeuAlaArgMetSerGlyAlaCysLeuValPro 239
720 CTGCTGGCAAGCGCTCGCTGGCGAACAAGTTTGATTTGACATCCGCC 769
239 evalProArgArgLysProAspGlyLysGlyTyrGlnLeuLeuMetLeu 256
770 CCGTCCAAAGGGGAATTGAAACGCGCAAAAGCCCAT...GATCGCGCGT 816
256 roProGluCysSerProProLeuAspAspAlaGluThrThrAlaAlaTr 272
817 TTCACCGCATGCCGAATATGATACGCCGTTTCCGACGACGATCT 866
273 MetAsnLysValValGluLysCysIleMetMetAlaProGluGlnTyr 289
867 GTTATGTACACCGCTACANA 888
289 tTrpLeuHisArgArgPheLys 296

seq_name: SwissProt_40:DDG_ECOLI

seq_documentation_block:
ID DDG_ECOLI STANDARD: PRT: 306 AA.
AC P76522: P76949.
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DDG protein.
GN DDG OR B2378.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / W3110;
RL Submitted (MAR-1997) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97349980; PubMed=9205837;
RA Yamamoto Y., Alba H., Baba T., Hayashi K., Inada T., Isono K.,

```

```

RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nishimoto H.,
RA Oshiba T., Oyama S., Saito N., Sempel G., Saton Y., Sivasubdaram S.,
RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,
RA Yamagata S., Horiuchi T.;
RT "Construction of a contiguous 874-kb sequence of the Escherichia coli
RT K-12 genome corresponding to 50.0-68.8 min on the linkage map and
RT analysis of its sequence features."
RL DNA Res. 4:91-113(1997).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (Potential).
CC -1- SIMILARITY: BELONGS TO THE HTRB/MSB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U49787; AAB6658.1; -
DR EMBL: AE000326; AAC75437.1; ALT_INIT.
DR EMBL: D90868; BAA16248.1; -.
DR EcoGene: EG12901; ddg.
KW Transmembrane; Inner membrane; Transferase; Acyltransferase;
KW Complete proteome.
FT TRANSMEM 20 40 POTENTIAL.
FT TRANSMEM 125 145 POTENTIAL.
SQ SEQUENCE 306 AA: 35493 MW: 58327F3D969E7B6 CRC64;

```

```

alignment_scores:
Quality: 189.50 Length: 294
Ratio: 1.199 Gaps: 12
Percent Similarity: 53.741 Percent Identity: 26.190

```

alignment_block:

US-09-303-518D-569 x DDG_ECOLI ..

Align seg 1/1 to: DDG_ECOLI from: 1 to: 306

```

88 CTGCGCGCTTCTGTCTGCACACGCTGGGAAACCGGCTGACATCTGCG 137
||||| ||| |||||||:||||| :|||
32 LeuProTyrProValLeuLysPheLysGlyThrArgIleGlyAlaMetAl 48
138 GTTATACCTTTAAAG..... 153
48 ArgProPheLeuLysArgArgGluSerIleAlaArgLysAsnLeuGluL 65
154 .....GAAGACCGCGCGCGCATCGTCCG... 177
65 euCysPheProGlnHisSerAlaGluGluArgGluLysMetIleAlaGlu 81
178 AATATCGCTGACGACGACATGAATCCGACCCCAAAAGCGTAAAGCCGT 227
|||||:||||| |||||
82 AsnPheArgSerLeuGlyMet..... 88
228 TTTTGGGAAAGCGCAAAAGCGGTTTGAACCTGCCCGCGGTTTTC 277
89 .....AlaLeuValGlnThrGlyMetAlaTrpPheT 99
278 GAAACCGGACATAGAACATATGTTCAAGCGGTACAGCGCGTGGGA 327
::: :||| |||:||||| :|
99 rProAspSerArgValArgLysTrpPheAsp...ValGluGlyLeuAsp 114
328 CATGTGACAGCGCTTGGACAAACAGAGGCTGATTCATCATCGCC 377
:|||||:||||| :|||
115 AsnLeuLysArgAlaGlnMetGlnAsnArgLysValMetValGlyIle 131
378 GCACATGCGGACATGATTTGGCGGACGCTACATCAGCCAGCGCTTC 427
||||| ||| :|||||:||||| :|||
131 HisPheMetSerLeuGluLeuGlyArgValMetGly.....Leuc 146

```

```

428 CGTCCCGCTGACCGCCATGTACAAACCGCCAAATCAACCATACAG 477
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
146 ysglnprmetmetalaThrtyrArgProHisAsnAsnGlnLeuMetIu 162
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
478 AAATCATGACGAGCGGAGGTTTCG...GGCAAGAGAAAACCGCGCC 524
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
163 TPVAlGlnThrArgGlyArgMetArgSerAsnLysAlaMetIleGlyAr 179
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
525 TACCAAGCATACAGGCGTCAACAAATCAATCAAGCCCTGCGTGGCGG 574
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
179 gsnAsnLeuArgGly.....IleValGlyAlaLeuLysGlyG 193
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
575 AAGCAACCATGCTCCGCCCGCACAGCTCCCTCCCTCAAGAAAGCGGG 624
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
193 lAlaValTyrPheAlaProasp.....GlnAspTyrGly 204
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
625 GAA.....GGCGTATGGGTGGATTCTTTCGCG...AAACCTGCTTATAC 665
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
205 ArgLysGlySerSerPheAlaProPhePheAlaValGluAsnValAlaTh 221
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
666 CATGACGCTGGCGGCAAAATTTGGCACAGCTCAAGCGCTGAAACCTGT 715
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
221 rHrsAsnGlyThrTyrValLeuSerArgLeuSerGlyAlaAlaMetLeuT 238
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
716 TTTTCTGCTGCGAGCGCTGCTGGCGGACAGTTTCGATTTCGCATC 765
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
238 hValThrMetValArgLysAlaAspTyrSerGlyTyrArgLeuThelle 254
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
766 CGCCCGCTCAAGGGCAATTTGACGCGC.....GACAAAGCCCATGA 806
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
255 ThrPro.....GluMetGluGlyTyrProThrAspGluAsnGlnAl 268
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
807 TGCGCGCGGTTCACACCGCATGCGCAATATGATACGCGCTTTCGCA 856
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
268 aAlaAlaTyrMetAsnLysIleIleGluLysGluIleMetArgAlaProG 285
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
857 CGCAGTATCTGTTATGTACAAACCGCTACAA 888
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
285 lGlnTyrLeuTyrPheIleHisArgArgPheLys 295
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
seq_name: SwissProt_40:MSBB_HAEIN
seq_documentation_block:
ID MSBB_HAEIN STANDARD; PRT; 318 AA.
AC P44567;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Lipid A biosynthesis (KDO)2-(lauroyl)-lipid IVA acyltransferase
DE (EC 2.3.1.-).
GN MSBB OR H10199.
OS Haemophilus influenzae.
OC Haemophilus: Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20 / ATCC 51907;
RC MEDLINE=95350630; PubMed=7542800;
RA Fletschmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fitchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RA "Whole-genome random sequencing and assembly of Haemophilus
RA influenzae Rd."
RL Science 269:496-512(1995).
-i- FUNCTION: TRANSFERS MYRISTATE OR LAURATE, ACTIVATED ON ACP, TO

```

```

CC (KDO)2-(LAUROYL)-LIPID IVA (BY SIMILARITY).
CC -i- PATHWAY: LIPOLYSACCHARIDE CORE BIOSYNTHESIS (BY SIMILARITY).
CC -i- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (Potential).
CC -i- SIMILARITY: BELONGS TO THE HTRB/MSBB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial/
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U32705; AAC21668.1; -.
DR TIGR: H10199; -.
KW Lipopolysaccharide biosynthesis; Transferase; Acyltransferase;
KW Transmembrane; Inner membrane; Complete proteome.
FT TRANSMEM 27 47 POTENTIAL.
FT TRANSMEM 91 111 POTENTIAL.
FT TRANSMEM 138 158 POTENTIAL.
SQ SEQUENCE 318 AA; 36882 MW; DE59952D/8719445 CRC64;

alignment_scores:
Quality: 178.50 Length: 275
Ratio: 1.137 Gaps: 9
Percent Similarity: 57.091 Percent Identity: 24.000

alignment_block:
US-09-303-518D-569 x MSBB_HAEIN ..
Align seg 1/1 to: MSBB_HAEIN from: 1 to: 318

70 CTCAATAGCTCTCCCTGCGCGCTTCTGTCACACGCTGGGAAA 119
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
39 LeuLeuLeuLeuAlaPheValProPheArgLeuAlaAspLysLeuThrG 55
120 CGGCGTGGACACATCGCGCTTTTACCTTTTAAAGAGAGCGCGCGCA 169
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
55 YLysLeuGlyYlIeTrrPleGlnHisLysAlaLysGlnArgThrArgA 72
170 TCGTCGCCAATATGCGTCAGCA.....GCCATGATCCCGAC 207
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
72 lAglnThrAsnLeuGlnTyrCysPheProHisTrrPheGluGlnGlnArg 88
208 CCCAAACGCTCAAGCCGTTTTCGCGAAACGCGAAAGCGGCTTGA 257
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
89 GluGlnValIleAspLysMetPheAlaValAlaGlnValMetPheG 105
258 ACTTGCCCCCGCGTTTTCAGAAAACCGGAAGACATGAAACAATGTCA 307
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
105 YlIeGlyGluIleAlaIleArgSerLysLysIleGlnLysArgSerG 122
308 AAGCGGTACACGCGCTGGGAACATGTGCAGCAGGCTTGACAAACGAA 357
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
122 lIurPheLle...GlyLeuGlnHisIleGluGlnAlaLysAlaGluGly 137
358 GGGCGTATTCATACGCGCGACATGCGGACATGATTTGGCGGAGC 407
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
138 AsnIleIleLeuMetValProHisGlyTrrAlaIleAspLaseArgYl 154
408 CTACATCAGCAGGACGCTTCGCTCCGCGTACGCGGACATGACAAACG 457
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
154 eIleLeuHisThrGln...GlyMetProMetThrSerMetLysAsnPro 170
458 CGAAATCAAAAGCGATGACAAATATGACAGCGGCGCGGCTTCGCGG 507
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
170 lSArgAsnProLeuValAspTrrPheThrIleThrArgGlnArgPhe 186
508 AAAGGAAAACCGCGCTTACAGCATCAAGGGGTCAACAAATCAAA 557
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
187 GlyGlyLysMetHisAlaArgGln...AsnGlyLysLysProPheLeu 202

```



```

31 AlaAlaMetAlaGlyLeuAlaLeuThrProProLysPheArgAspProI 47
114 GGGAAACCGCGTCGACATCGCGCTTTACCTTTAAAGAAAGCCGG 163
47 eLeuAlaIaGlyLeuGlyArgPheAlaIaGlyLeuGlyLysSerSerArg 64
164 CGCGCATCGTCGCCAATATG..... 183
64 rgrAlaLeuLeuLeuLeuSerLeuCysPheProGluArgSerGluAla 80
184 ...CGTCAGGAGCGATGATCCCGACCCCAAAACGGTCAAAAGCGTTT 230
81 GluArgGluAla.....IleValAspGluMetPh 90
231 TCGGGAACGCGCAAAAGCGGTTTGAACTTCGCCCGCGGTTTTCAGAA 280
90 eAlaThrAlaProGlnAlaMetAlaMetLeuAlaIaLeuAlaIaArg 107
281 AACCGAAGACATAGAACAAATGTTCAAGCGGTACAGCGGTGGAGACAT 330
107 LProGluLysIleGlnProArgValAsp...TrpGlnGlyLeuGluIle 122
331 GGGCAGGAGCGGTTTGACAAACACAGAGCGGTGCTATTTCATCAGCGCA 380
123 ILeGluGluMetArgArgAsnAsnGluLysValIlePheLeuValProH 139
381 CATCGGACGCTACGATTGGCGGAGCGGTACATCAGCAGCGGTCGCT 430
139 seLysTrpAlaValAspIleProAlaMetLeuMetAlaSerGln...Gly 155
431 TCCCGCTGACCGCCATGTACAAACCGCGAAATCAAGCGATAGACAAA 480
155 InLysMetAlaIaMetPheHisAsnGlnGlyAsnProValPheSerPyr 171
481 ATCATGACGAGCGGCGGCTTCGCGCAAGAAAGAACCGCGCTACAG 530
172 ValTrpAsnThrValAlaArgArgPheGlyLysArgLeuHisAlaValAs 188
531 CATACAGGCGGTCAAAACAAATCATCAAAAGCGGTGCGGCGCAAGCAA 580
188 n...AspGlyIleLysProPheIleGlnSerValArgGlnGlyTrpG 204
581 CCATCGTCTGCTGCC....GACCAAGTCCCTCCCTCAAGAAAGCGGG 624
204 LysTrpLysLeuProAspGlnAspHisGlyProGlnHisSerGlu..... 218
625 GAAGCGGTATGGGTATCTTCGCGCAACCTGCTATACCATGACGCT 674
219 .....PheValAspPhePheAlaThrTrpLysAlaThrLeuProAl 232
675 GCGGCAAAATTTGGCACACGTCAAAGGCGGTGAAACCTGTTTTCGCT 724
232 aIleGlyArgLeuMetLysVal.....CysA 241
725 GCGAAGCGCTG.....CCTGGCGGACAAAGT..... 750
241 rGAlaIaGlyValAlaProLeuPheProIleTyrAspGlyLysThrHisArg 257
751 TTGATTTGCACATCGCGCCGCTCCAGGGGAA...TTGAAGCGGACAA 797
258 LeuThrIleGlnValArgProPheMetAspAspLeuLeuGluAlaAspAs 274
798 AGCCCATGATGCCCGCTGTTCAACCGCAATCGGAATATTGATGACGC 847
274 PhisThrIleAlaArgArgMetAsnGluGluValGluIlePheValGly 291
848 GTTTTCGACGACATATCTGTTTATGTACAAACCGCTACAA 888
291 roArgProGluGlnIleTrpThrIleLeuLysLeuLeuLys 304
seq_name: SwissProt_40:YG3_HALSQ
seq_documentation_block:

```

```

ID YG3_HALSQ STANDARD; PRT; 437 AA.
AC P21561;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 50.6 kDa protein in the 5' region of GYRA and GYRB (ORF
DE 3).
OS Haloferax sp. (strain Aa 2.2).
OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae; Haloferax.
OX NCBI_TaxID=2254;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91100352; PubMed=1846146;
RA Holmes M.L., Dyall-Smith M.L.;
RT "Mutations in DNA gyrase result in novobiocin resistance in
RT halophilic archaeobacteria."
RL J. Bacteriol. 173:642-648(1991).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb-sdb.ch).
CC -----
CC EMBL: M3873; -; NOT_ANNOTATED_CDS.
DR PIR: C39135; C39135.
KW Hypothetical protein.
SQ SEQUENCE 437 AA; 50626 MW; B5B99A2AF3892BEF CRC64;

```

alignment_scores: Quality: 115.00 Length: 296
Ratio: 0.935 Gaps: 16
Percent Similarity: 41.554 Percent Identity: 26.014

alignment_block:

US-09-303-518D-569 x YG3_HALSQ ..

Align seq 1/1 to: YG3_HALSQ from: 1 to: 437

```

108 CAGCGTCGGAACCGCGTCGACATCGCGGCTTTACCTTTAAAGGANG 157
119 HisAlaGly...AspArgAlaArgProGlyVal.....AspSerArg 131
158 ACCGCGCGCGCATGTCGCCAATATGCTGACGAGCATGAATCC...C 204
131 gLeuArg...GlnGlnHisGlnHisProArgGlyArgHisAlaSerAspA 147
205 GACCCCAAAACGGT.....CAAGC 224
147 rGAlaIaGlnAspGlyAlaHisProAlaArgGlnArgLeuArgGluGlnPro 163
225 CGTTTTTCGGAACCGCGCAAGGCGGTTTGAACTTCGCCCGCGCTTT 274
164 ArgHisAlaGlyArgProArgArgArg.....GlnProProArgArgAl 178
275 TCGAAACCGGACGATAGAAACAAATGTTCAAAAGCGGTACAGCGCTG 324
178 yArgSerArgGlyThrHisArgArgHisLeuArgGlnAlaProAlaPro 195
325 GAACATGTCAGCAGCGCTTTGGCAACACAGAAAGGCGTCTATTGATCAG 374
195 lValArgGlyProAspGlnAspGlnAlaArgGlu..... 206
375 GCGGACATCGCGACGCTACGATTTGGCGGAGCGCTACATCAGCGACAG 424
207 .....PheArgGlyProArgHisArgArgGluArgHisProProThrAl 221
425 TTCGGTTCCCGCGTACGCGCATGACAAACCGCGCAAAATCAAGCGATA 474
221 a.....ArgAspValLeuArgGlyGluProGlnHisGlyA 233

```

```

475 GACAAATCATGACGCGGCGAGGTTCCGGGCAAGGAAAAACCGGCC 524
      :::::  |||
233 spg1h1sh1sh1eug1ug1yArg..... 240
525 TACAGACATCAAGGGTCAAAATCATCAAAAGCCCTGCTGGGCG 574
      |||||  ::  |||
241 .....Arg1yArgProArgProGlnGlyArgGlnAlaGlyArg 253
      |||||  ::  |||
575 AAGCAACCATGCTCT..... 590
      |||  |||
253 gGlyAlaHisProGlnValArgAlaArg1yLeuAlaAlaGlyG 270
      |||  |||
591 .....GCCGACCAAGTCCCTCCCTCAAGA..... AGG 620
      :::::  |||||
270 LuAlaArg1yLeuProGlnProArgProLeuGlyValArgThValHis 286
      |||  |||
621 CGGGAGAGCGGTATGGGTGATTTCTTCGCAACCTGCTATACATGA 670
      |||||  |||  ::  |||
287 ArgGlyArg1yLeuArgGlyArgValGlyAlaGlyProArgProG 303
      ::  |||  ::  |||
671 CGCTGGCGGCAAAATGGCACACGT.....CAAAAGCGTGAAA 708
      ::  |||  ::  |||
303 nValProGlyAspPheAlaProGlnGlyLysPserGlnArgGlyL 320
      ::  |||  ::  |||
709 ACCCGTTTCTGCTGCGAAGCGCTGCC..... 737
      |||  |||
320 hProPro.....ProArgProHisSerArgLysArgArg 330
      |||  |||
738 .....TGCGGACAAAGTTTCA..... 755
      |||  |||
331 AspThrGlyAlaHisHisArgHisThrArgArgArgArgValArg 347
      |||  |||
756 .....TTTGCACATCCCGCCCGTCCAGAGGGAA 783
      |||  |||
347 gHisArgGlnGlyAlaLeuProAlaAlaHisProAspArgArgArg 364
      |||  |||
784 TTGAACGGCGCAAAAGCCCATGATCCCGCTGTTC 819
      :::::  |||  |||  |||  :::::
364 rGArgArgArgAlaHisPro.AsPheAlaAlaLysr 375
      :::::  |||  |||  |||  :::::

seq_name: SwissProt_40:MSBB_SHIFL

seq_documentation_block:
ID MSBB_SHIFL STANDARD; PRT; 231 AA.
AC 006659:
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Lipid A biosynthesis (KDO2-(lauroyl))-lipid IVA acyltransferase
DE (EC 2.3.1.-) (Fragment).
OS MSBB.
OS Shigella flexneri.
OC Plasmid 230 kb pMSH6000.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Shigella.
OX NCBI_TaxID=623;
OX [1]
RN SEQUENCE FROM N.A.
RC STRAIN=2a.
RX MEDLINE=97315241; PubMed=9171415;
RA Radnedge L., Davis M.A., Youngren S.J.;
RT "Plasmid maintenance functions of the large virulence plasmid of
RT Shigella flexneri.";
RL J. Bacteriol. 179:3670-3675(1997).
CC -1- FUNCTION: TRANSFERS MYRISTATE OR LAURATE, ACTIVATED ON ACP, TO
CC (KDO)2-(LAUROYL)-LIPID IVA (BY SIMILARITY).
CC -1- PATHWAY: LIPIDOLYSACCHARIDE CORE BIOSYNTHESIS (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (potential).
CC -1- SIMILARITY: BELONGS TO THE HTRB/MSBB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U82621; AAB58154.1;
CC KW Lipopolysaccharide biosynthesis; Transferase; Acyltransferase;
KW Transmembrane; Inner membrane; Plasmid.
FT NON_TER 1 1
FT TRANSMEM 48 68 POTENTIAL
FT SEQUENCE 231 AA; 26754 MW; 99350DB0A499DB54 CRC64;
SQ

```

```

alignment_scores:
  Quality: 112.00      Length: 245
  Ratio: 0.933        Gaps: 8
  Percent Similarity: 48.980      Percent Identity: 21.224

```

```

alignment_block:
US-09-303-518D-569 x MSBB_SHIFL ..

```

```

Align seg 1/1 to: MSBB_SHIFL from: 1 to: 231

```

```

217 GTCAAGCCGTTTGGCGAAAGCGCAAAAGCGGTTTGGAACTTGCCCC 266
    |||  :::::  ::  :::::
1 ValAspAsnMetPheAlaThrAlaLeuGlnSer1LeValMetMetAlaG 17
267 CGCGTTTTCAGAAACCGGAGACATAGAAACAATGTCAAGCGGTA. 315
    |||  :::::  |||  |||
17 uLeuAla1LeuArg1yProGlnLysPheGln.....LysArgValr 31
316 .....CACGGCTGGGACACATGTGCACAGCGCTTGGACAAACAGAAAGCG 360
    |||  |||  :::::
31 hettPrysg1yLeuGln1LeuGlnGln1u1eArgHisAsnAsnArgAsn 47
361 CTCTCTATTCATACGCGCCGACATCGGACGTACGATTTGGCGGAGCGTA 410
    :::::  |||  :::::  |||  :::::
48 Val1LePheLeuValrProHisLysLysPserValAsp1LeProAlaMet 64
411 CATCAGCGCAGCAGCTTCCTCCGCTCCGCGCAGCATGACAAACCGCGCA 460
    :::::  |||  :::::  |||  :::::
64 uLeuAla1AlaGln...GlyGlnLysMetAlaAlaMetPheHisGlnGln 80
461 AAATCAAGCGATAGACAAATATCATGCAGCGGCGAGGTTGCCGCAAA 510
    ::  :::::  |||  ::  |||
80 rGAsnProVal1LeAsp1yValr1PAsnSerValArg...ArgLysPhe 95
511 GGAAGAAACCGCGCTACGACATACAAAGGGGTCAAAACATCATCAAGC 560
    |||  ::  |||  :::::  |||  :::::
96 GlyGlyArgLeuHisSerArgGlnAspGly1LeLysProPhe1LeGln 112
561 CCGCGCTTGGCGGAGCAACATCATGCTCTGCCC.....GACCAGCTCC 604
    :::::  |||  |||||  |||||
112 rValArgGlnGly1yTr1Pry1yTr1yTr1yLeuProAspGlnAspHisGlyr 129
605 CCTCCCTCAAGAGCGCGGGAAGCGGTATGGTGGATTTCTTGGCAAA 654
    |||  :::::  |||  :::::  |||  :::::
129 rGln1yTrSerGln.....PheAlaAspPheAlaThr 140
655 CTGCGCTATACATGACGCTGGGCGCAAAATGGCACAGCTCAAGCGGT 704
    |||  ::  |||  :::::  |||  :::::
141 TyrLysAlaThrLeuPro1Le1LeGlyArgLeuMetAsn1Le..... 154
705 GAAGAACCTGTTTCTGTGCGAAGCGCTGCTGGCGGCAAGAGTTTGC 754
    :::::  |||  |||  :::::
155 .....SerGlnAlaMet 159
755 ATTTGCACATCCGCCCGTCAAGGGGA..... 783
    ::  ::  |||||  |||
159 1e1LeProLeuPheProVal1yTrAspGlnLysLysHisPheLeuThr1Le 175

```


784TTGAACGGCCGCAAGCCCA 803
 176 GluValAlaGProPromeMetaspalacysIIleAlaSerIalaspnslySke 192
 804 TGATGCCCGCGCTTTCACACCGCATGCCAATATTGGATACGCCGCTTTTC 853
 192 ttleAlaArgIgmelaSlnlyStrValIgluIleuValIglYserHisp 209
 854 CGACGACATCTCTTATGTACACCGCTACAA 888
 209 rogluGlnTyrIleTrrpValleuIleuIleuIlys 220

seq_name: SwissProt_40:NFX1_HUMAN

seq_documentation_block:
 ID NFX1_HUMAN STANDARD; PRT; 1104 AA.
 AC 012986;

DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Transcriptional repressor NF-X1 (Nuclear transcription factor, X box-
 binding, 1).
 GN NFX1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OX NCBI_Taxid=9606;

RP SEQUENCE FROM N. A.
 RX MEDLINE=95053707; PubMed=7964459;
 RA Song Z., Krishna S., Thanos D., Strominger J.L., Ono S.J.;

RT "A novel cysteine-rich sequence-specific DNA-binding protein
 RT interacts with the conserved X-box motif of the human major
 RT histocompatibility complex class II genes via a repeated Cys-His
 RT domain and functions as a transcriptional repressor.";
 RL J. Exp. Med. 180:1163-1174 (1994).
 CC -1- FUNCTION: REPRESSOR OF HLA-DRA TRANSCRIPTION. BINDS TO THE X-BOX
 CC MOTIF OF CLASS II MHC GENES. MAY PLAY AN IMPORTANT ROLE IN
 CC REGULATING THE DURATION OF AN INFLAMMATORY RESPONSE BY LIMITING
 CC THE PERIOD IN WHICH CLASS II MHC MOLECULES ARE INDUCED BY
 CC INTERFERON-GAMMA.

CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- INDUCTION: BY INTERFERON GAMMA.
 CC -1- SIMILARITY: TO D.MELANOGASTER SHUTTLE CRAFT PROTEIN (STC) AND
 CC YEAST YND23C.

CC -1- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@sdb-sdb.ch).

CC EMBL: U15306; AAA69517.1; -;
 DR HSP: P02876; 2WC.
 DR MIM: 603255; -;
 DR InterPro: IPR001965; PHD.
 DR InterPro: IPR001374; R3H.
 DR InterPro: IPR000967; ZNF_NFX1.
 DR InterPro: IPR001841; Znf_finger.
 DR Pfam: PF01424; R3H; 1.
 DR Pfam: PF01422; ZF-NFX1; 8.
 DR SMART: SM00393; R3H; 1.
 DR SMART: SM00184; RING; 1.
 DR SMART: SM00438; ZNF_NFX1; 9.
 DR PROSITE: PS00518; ZF_RING_1; FALSE_NBS.
 DR PROSITE: PS00089; ZF_RING_2; 1.
 DR Transcription regulation; Repressor; DNA-binding; Nuclear protein;
 KW Repeat; Zinc-finger.

FT 342 393 RING-TYPE.
 FT ZN-FING 424 876 7 X APPROXIMATE REPEATS, CYS-RICH.
 FT DOMAIN

FT REPEAT 424 463 1.
 FT REPEAT 480 517 2.
 FT REPEAT 541 580 3.
 FT REPEAT 605 647 4.
 FT REPEAT 695 734 5.
 FT REPEAT 806 842 6.
 FT REPEAT 843 876 7.
 SQ SEQUENCE 1104 AA; 123149 MW; 36FEC4292F78130C CRC64;

alignment_scores:

Quality: 112.00 Length: 368
 Ratio: 0.848 Gaps: 18
 Percent Similarity: 35.870 Percent Identity: 20.380

alignment_block:

US-09-303-518D-569 x NFX1_HUMAN

Align seg 1/1 to: NFX1_HUMAN from: 1 to: 1104

14 AATTGACGCTGTTTCCCTTTCGACACG...CCATGCACATCGT... 58
 ||| ||| ::::: ||| ||| ||| |||
 553 AsnHisThrCysSerGlnValCysHisProGlnProCysGlnGlnCysPr 569
 59TGACCGCCCTGCTCAATGCTCTCCCTGCTGCGCTTCTGCT 104
 ||| ||| ::::: ||| ||| ||| |||
 569 oArgLeuProGlnLeuValArgCysCysProCysGlyGlnThrProLeus 586
 105 GCACAGCGCTGGGAAACCGGCTCGACATCTGCGTTTACCTTTAAAG 154
 ||| ||| ||| ||| ||| ||| ||| |||
 586 ergInleuLeuGlnleuGlnGlnSerSer.....Arg 596
 155 AATACCGCGCGCGCATCTGCGCAATATGCTGACGACGACATGATCC 204
 ||| ||| ||| ||| ||| ||| ||| |||
 597 LysThrCysMetaspProValProSerCysGlyValCysGlyLysPr 613
 205 GACCCCAAAACGTCAAACCGCTTTTGGGAAACGGCAAAAGCGGTT 254
 ||| ||| ||| ||| ||| ||| ||| |||
 613 oleuProCysGlnGlnSerLeuAspPheIle..... 622
 255 GGAACCTGCCCGCGGCTTTTCAGAAACCGGAACATAGAACAAATGT 304
 622 622
 305 TCAAAGCGGTACAGCGCTGGGAAACATGTCAGACAGCGCTTTGACAAAC 354
 ::::: ||| ||| ||| ||| ||| ||| |||
 623HisThrCysGlnLeuLeuCys.....HisG 631
 355 GAAGGCTGCTATTTCATCAGCGCGCATCGCACTACGATTGGGCGG 404
 ||| ||| ||| ||| ||| ||| ||| |||
 631 uGlyAspCysGlyProValSerArgThrSerVal..... 642
 405 AGCTACATCAGCGCAGCGCTTCCGCTCCGCTACCGCATGTACAAAC 454
 ||| ||| ||| ||| ||| ||| ||| |||
 643 ..IleSerCysArgCysSerPheArgThrLysGlnLeuProCysThrSer 658
 455 CGCCGAAATCAAGCGCATAGACA.....AAATCATGACGAGCG 492
 ||| ||| ||| ||| ||| ||| ||| |||
 659 ..LeuLysSerGlnAlaPalaThrPheMetCysAspLysArgCysAsnly 674
 493 GCGAGGCTTCGCGGCAAGGAAANA..... 517
 674 sLysArgLeuCysGlnArgHisLysCysAsnGlnIleCysCysValAsp 691
 518CCGCGCTACCGACATACAG.....GGTCA 544
 691 ysgLHisLysCysProLeuAsnCysGlyArgLysLeuArgCysGlyLeu 707
 545 AACCAATCATCAAGCCCTGCTGGGCGAAG..... 577
 708 HisArgCysGlnGlnProCysHisArgGlyAsnCysGlnThrCysTrp 724
 578C 578


```

798 1stYrGlycInThrGlyGlyGlyLeuSerSerHisGlnGlyInThr 814
501 TCAGGGGCAAGAAACCGCGCTACGACATACA.....AGGGGTCA 544
    :||| :||| :||| :||| :||| :||| :||| :|||
815 AspSerGlnGlyGlnAsnSerGlnInThrPHisArgThrAspSerGlnGly 831
545 AACCAATCAT.....CAAGCCCTCGCTCGGCGGCAACGACATCGTC 568
    :||| :||| :||| :||| :||| :||| :||| :|||
831 nserPHeHisPheAspGlnAlaGlyArgGlyGlnSerSerHisHisG 848
589 CTGGCCGACACAGTCCTCCCTCCCTCAAGAGCGGGGAGG...CGATG 635
    :||| :||| :||| :||| :||| :||| :||| :|||
848 LysInThrAspArgGlnSerGlnSerHisGlyGlnSerGlnGly 864
636 GGTGATTTCTTCGCAACCTCGCTTACCATGACGCTGC..... 677
    :||| :||| :||| :||| :||| :||| :||| :|||
865 GlyThrGlnAsnGlnGlnGlnGlnAsnArgHisSerLeuGlyThrAsp 881
678 .....GGCAAAATG 687
881 gThrArgArgAspSerTyValGlnHisSerGlyArgSerGlyLysLeu 898
    :||| :||| :||| :||| :||| :||| :||| :|||
898 ergInGlnAsnSerArgGlnGlu.....Val 906
738 TGGCGGACAGAGTTTCATTTGCACATCCGCCCTCCAGGGAATGA 787
    :||| :||| :||| :||| :||| :||| :||| :|||
907 ArgInThrGlnSerGlnArgSerHisAspArgArgGlnGlnGlnIle 923
788 ACGGCGA 794
923 nGlnGln 925

seq_name: SwissProt_40:AT7A_HUMAN

seq_documentation_block:
ID AT7A_HUMAN STANDARD; PRT; 1500 AA.
AC 004656; 000745; 000227;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Copper-transporting ATPase 1 (EC 3.6.3.4) (Copper pump 1) (Menkes
DE disease-associated protein).
GN ATP7A OR MNK OR MCL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 4).
RC TISSUE=Fibroblast;
RX MEDLINE=93258410; PubMed=8490659;
RA Vulpe C.D., Levinson B., Whitney S., Packman S., Gitschier J.;
RT "Isolation of a candidate gene for Menkes disease and evidence that
RT it encodes a copper-transporting ATPase.";
RL Nat. Genet. 3:7-13(1993).
RN [2]
RP ERRATUM.
RA Vulpe C.D., Levinson B., Whitney S., Packman S., Gitschier J.;
RL Nat. Genet. 3:273-273(1993).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 4).
RX MEDLINE=9533177; PubMed=7607665;
RA Tuemer Z., Vural B., Toennesen T., Chelly J., Monaco A.P., Horn N.;
RT "Characterization of the exon structure of the Menkes disease gene
RT using vectorite PCR.";
RL Genomics 26:437-442(1995).
RN [4]
RP SEQUENCE OF 1-1447 FROM N.A. (ISOFORM 4).
RX MEDLINE=96039257; PubMed=7490081;
RA Dietrick H.A., Ambrosini L., Spencer J., Glover T.W., Mercer J.F.B.;
RT "Molecular structure of the Menkes disease gene (ATP7A).";

```

```

RL Genomics 28:462-469(1995).
RN [5]
RP SEQUENCE OF 1-626 FROM N.A. (ISOFORM 4).
RC TISSUE=Kidney;
RX MEDLINE=93258397; PubMed=8490646;
RA Chelly J., Tuemer Z., Toennesen T., Petterson A., Ishikawa-Brush Y.,
RA Tommerup N., Horn N., Monaco A.P.;
RT "Isolation of a candidate gene for Menkes disease that encodes a
RT potential heavy metal binding protein.";
RL Nat. Genet. 3:14-19(1993).
RN [6]
RP SEQUENCE OF 12-529 FROM N.A. (ISOFORM 4).
RC TISSUE=Endothelial cells;
RX MEDLINE=93258398; PubMed=8490647;
RA Mercer J.F.B., Livingston J., Hall B., Paynter J.A., Begy C.,
RA Chandrasekharappa S., Lockhart P., Grimes A., Bhave M.,
RA Stenmark D., Glover T.W.;
RT "Isolation of a partial candidate gene for Menkes disease by
RT positional cloning.";
RL Nat. Genet. 3:20-25(1993).
RN [7]
RP SEQUENCE FROM N.A. (ISOFORM 4).
RA Bird C.;
RL Submitted (APR-1997) to the EMBL/Genbank/DBJ databases.
RN [8]
RP SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).
RC TISSUE=Fibroblast, and Colon carcinoma;
RX MEDLINE=99179550; PubMed=10079814;
RA Harris E.D., Reddy M.C., Qian Y., Tiffany-Castiglioni E., Majumdar S.,
RA Nelson J.;
RT "Multiple forms of the Menkes Cu-ATPase.";
RL Adv. Exp. Med. Biol. 448:39-51(1999).
RN [9]
RP SEQUENCE FROM N.A. (ISOFORM 3).
RC TISSUE=Fibroblast;
RX MEDLINE=98359731; PubMed=9693104;
RA Reddy M.C., Harris E.D.;
RT "Alternative splicing of Menkes mRNA.";
RL Biochem. J. 334:71-77(1998).
RN [10]
RP ALTERNATIVE SPLICING (ISOFORM 5), AND SUBCELLULAR LOCATION.
RX MEDLINE=98133927; PubMed=9467005;
RA Qi M., Byers P.H.;
RT "Constitutive skipping of alternatively spliced exon 10 in the ATP7A
RT gene abolishes Golgi localization of the menkes protein and produces
RT the occipital horn syndrome.";
RL Hum. Mol. Genet. 7:465-469(1998).
RN [11]
RP ALTERNATIVE SPLICING (ISOFORM 6).
RC TISSUE=Neuroblastoma;
RX MEDLINE=20427712; PubMed=10970802;
RA Reddy M.C., Majumdar S., Harris E.D.;
RT "Evidence for a menkes-like protein with a nuclear targeting
RT sequence.";
RL Biochem. J. 350:855-863(2000).
RN [12]
RP SUBCELLULAR LOCATION.
RX MEDLINE=97227286; PubMed=9147644;
RA Dietrick H.A., Adam A.N., Escara-Wilke J.F., Glover T.W.;
RT "Immunocytochemical localization of the Menkes copper transport
RT protein (ATP7A) to the trans-Golgi network.";
RL Hum. Mol. Genet. 6:409-416(1997).
RN [13]
RP SUBCELLULAR LOCATION, AND MUTAGENESIS OF LEUCINE RESIDUES.
RX MEDLINE=99415752; PubMed=10484781;
RA Petris M.J., Mercer J.F.;
RT "The Menkes protein (ATP7A; MNK) cycles via the plasma membrane both
RT in basal and elevated extracellular copper using a C-terminal
RT di-leucine endocytic signal.";
RL Hum. Mol. Genet. 8:2107-2115(1999).
RN [14]
RP STRUCTURE BY NMR OF 375-446.

```

RA MEDLINE-9810082; PubMed-9437429;
 RA Gitschier J., Moffat B., Reilly D., Fairbrother W.J.,
 RT "Solution structure of the fourth metal-binding domain from the
 RT Menkes copper-transporting ATPase.";
 RL Nat. Struct. Biol. 5:47-54(1998).
 RN [15]
 RP REVIEW, AND VARIANTS MD.
 RX MEDLINE-9179553; PubMed-10079817;
 RA Tuemer Z., Moeller L.B., Horn N.,
 RT "Mutation spectrum of ATP7A, the gene defective in Menkes disease.";
 RL Adv. Exp. Med. Biol. 448:83-95(1999).
 RN [16]
 RP VARIANT LEU-767, AND VARIANT MD ARG-1302.
 RX MEDLINE-95067953; PubMed-7977350;
 RA Das S., Levinson B., Whitney S., Vulpe C., Packman S., Gitschier J.,
 RT "Diverse mutations in patients with Menkes disease often lead to exon
 RT skipping.";
 RL Am. J. Hum. Genet. 55:883-889(1994).
 RN [17]
 RP VARIANTS MD PRO-629; ARG-727; PRO-1006 AND ASP-1019.
 RX MEDLINE-97136571; PubMed-8981948;
 RA Tuemer Z., Lund C., Tolshave J., Vural B., Toennesen T., Horn N.,
 RT "Identification of point mutations in 41 unrelated patients affected
 RT with Menkes disease.";
 RL Am. J. Hum. Genet. 60:63-71(1997).
 RN [18]
 RP VARIANT OHS LEU-637.
 RX MEDLINE-97388773; PubMed-9246006;
 RA Ronce N., Moizard M.P., Robb L., Toulain A., Villard L., Moraine C.,
 RT "A C205T transition in exon 8 of the ATP7A gene is associated with
 RT exon skipping in an occipital horn syndrome family.";
 RL Am. J. Hum. Genet. 61:233-238(1997).
 RN [19]
 RP VARIANT MD VAL-1362.
 RX MEDLINE-99330564; PubMed-10401004;
 RA Ambrosini L., Mercer J.F.,
 RT "Defective copper-induced trafficking and localization of the Menkes
 RT protein in patients with mild and copper-treated classical Menkes
 RT disease.";
 RL Hum. Mol. Genet. 8:1547-1555(1999).
 RN [20]
 RP VARIANT MD ARG-873.
 RX MEDLINE-99253147; PubMed-10319589;
 RA Ogawa A., Yamamoto S., Takayanagi M., Kogo T., Kanazawa M., Kohno Y.,
 RT "Identification of three novel mutations in the MNK gene in three
 RT unrelated Japanese patients with classical Menkes disease.";
 RL J. Hum. Genet. 44:206-209(1999).
 RN [21]
 RP FUNCTION: MAY SUPPLY COPPER TO COPPER-REQUIRING PROTEINS WITHIN
 CC THE SECRETORY PATHWAY, WHEN LOCALIZED IN THE TRANS-GOLGI NETWORK,
 CC UNDER CONDITIONS OF ELEVATED EXTRACELLULAR COPPER, IT RELOCALIZED
 CC TO THE PLASMA MEMBRANE WHERE IT FUNCTIONS IN THE EFFLUX OF COPPER
 CC FROM CELLS.
 CC [22]
 RP CATALYTIC ACTIVITY: ATP + H(2)O + CU(2+)(IN) = ADP + PHOSPHATE +
 CC CU(2+)(OUT).
 CC [23]
 RP SUBUNIT: MONOMER.
 CC [24]
 RP SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. CYCLES
 CC CONSTITUTIVELY BETWEEN THE TRANS-GOLGI NETWORK (TGN) AND THE
 CC PLASMA MEMBRANE. PREDOMINANTLY FOUND IN THE TGN AND RELOCALIZED TO
 CC THE PLASMA MEMBRANE IN RESPONSE TO ELEVATED COPPER LEVELS. ISOFORM
 CC 3 MAY BE CYTOSOLIC. ISOFORM 5 IS LOCATED IN THE ENDOPLASMIC
 CC RETICULUM.
 CC [25]
 RP ALTERNATIVE PRODUCTS: 6 ISOFORMS: 1, 2, 3/2-16, 4 (SHOWN HERE), 5
 CC AND 6/NL45; ARE PRODUCED BY ALTERNATIVE SPLICING. ISOFORM 5 LACKS
 CC THE TRANSMEMBRANE DOMAINS 3 AND 4, IT IS EXPRESSED AT A LOW LEVEL
 CC IN SEVERAL TISSUES OF NORMAL INDIVIDUALS AND IS THE ONLY ISOFORM
 CC FOUND IN PATIENTS WITH OHS. ISOFORM 3/2-16 LACKS 6 TRANSMEMBRANE
 CC REGIONS AND 5 HEAVY-METAL-ASSOCIATED (HMA) DOMAINS. ISOFORM 6
 CC LACKS ALL TRANSMEMBRANE REGIONS AND 5 HEAVY-METAL-ASSOCIATED (HMA)
 CC DOMAINS, BUT HAS A PUTATIVE NUCLEAR LOCALIZATION SIGNAL ATTACHED
 CC AT THE N-TERMINUS.
 CC [26]
 RP TISSUE SPECIFICITY: FOUND IN MOST TISSUES EXCEPT LIVER. ISOFORM
 CC 3 IS WIDELY EXPRESSED INCLUDING IN LIVER CELL LINES. ISOFORM
 CC 1 IS EXPRESSED IN FIBROBLASTS, CHORIOCARCINOMA, COLON CARCINOMA

CC AND NEUROBLASTOMA CELL LINES. ISOFORM 2 IS EXPRESSED IN
 CC FIBROBLASTS, COLON CARCINOMA AND NEUROBLASTOMA CELL LINES.
 CC [27]
 RP DISEASE: DEFECTS IN ATP7A ARE ASSOCIATED WITH MENKES SYNDROME (MD)
 CC (ALSO KNOWN AS KINKY HAIR DISEASE). AN X-LINKED RECESSIVE DISEASE
 CC CHARACTERIZED BY PROGRESSIVE NEURODEGENERATION AND CONNECTIVE-
 CC TISSUE DISTURBANCES: FOCAL CEREBRAL AND CEREBELLAR DEGENERATION,
 CC EARLY RETARDATION IN GROWTH, PECULIAR HAIR, HIPPODYSPLASIA,
 CC CUTIS LAXA, VASCULAR COMPLICATIONS AND DEATH IN EARLY CHILDHOOD.
 CC IT IS DUE TO A DEFECT IN ABSORPTION AND TRANSPORT OF COPPER.
 CC [28]
 RP DISEASE: DEFECTS IN ATP7A ARE ALSO ASSOCIATED WITH OCCIPITAL
 CC HORN SYNDROME (OHS) (ALSO KNOWN AS X-LINKED CUTIS LAXA).
 CC [29]
 RP MISCELLANEOUS: THE C-TERMINAL DI-LEUCINE, LEU-1487 ET LEU-1488, IS
 CC AN ENDOCYTIC TARGETING SIGNAL WHICH FUNCTIONS IN RETRIEVING
 CC RECYCLING FROM THE PLASMA MEMBRANE TO THE TGN. MUTATION OF THE DI-
 CC LEUCINE SIGNAL RESULTS IN THE ACCUMULATION OF THE PROTEIN IN THE
 CC PLASMA MEMBRANE.
 CC [30]
 RP SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY
 CC (E1-E2 ATPASES). SUBFAMILY IB.

alignment_scores:
 Quality: 104.50 Length: 232
 Ratio: 0.823 Gaps: 15
 Percent Similarity: 54.741 Percent Identity: 25.862

alignment_block:

US-09-303-518d-569 x AT7A_HUMAN ..

Align seg 1/1 to: AT7A_HUMAN from: 1 to: 1500

97 TCCTGTGTGACACGCTGGAAACCGGCTGGACA 131
 21 SerCysValITPrHrIleGluGlnGlnIleGlyValValnGlyValHl 37
 132 TCCTGGCGTTTACCTTTAAAGAACCGCGCGCATGTCGCGCA 181
 37 shIleValValserLeuGluGlnGlnValAsnAlaThrIleIleIleYrAsp 54
 182 TGGCTGACGCGCATGATATCCGACCCCAAAAGCGTCAAGCGGTTT 231
 54 rOlYsLeuGlnThrProLYsThrLeuGlnGlnAlaIle 66
 232 GCGGAACGCGCAAGCGGTTTGAACCTGCCCCCGCTTTTCGA 281
 67 AspAspMetGlyPheAspAlaValIleHis 77
 282 ACCGGAAGACATAGAACATGTTCAAGCGGTACAC 319
 77 nProAspProLeuProValLeuThrAspThrLeuPheLeuThrValThr 94
 320 GCTGGACATGTGACGACGCGTTTGGCAAAAC 354
 94 laserLeuThrLeuProTrpAspHisIleGlnserThrLeuLeuLYs 110
 355 GAAGGCTGTATTCATGACCGCGACATCGCGCAG 389
 110 rLYsGlyValThrAspIleLYsIleLYrProGlnLYsArgThrValAlaY 127
 390 CTACAGATTGGCGGACGATACATGACGACGCGTTCGTTCCGCTCA 439
 127 alhrlleIleProserIleValAsnAlaValnIleLYsGluLeuVal 143
 440 CGCCCATGTACAAACCGCGAAATCAAC 470
 144 Pro.GluLeuSerLeuAspThrGlyThrLeuGluLYsSerGlyAlaC 160
 471GATGACAAATCATGACGCGGCGCAGGTTCG 503
 160 ysluAspHis.SerMetAlaGlnAlaGlyGluValValLeuLYsmety 176
 504 CGGCAAGGAAAGAAACCGCGCTAC 528
 176 sValGluGlyMetThrCysHisSerCysThrSerThrIleGluGlyLYsI 193

```

529 .....AGCATACAGGGGTCAACAATCATCAAGCCCTGCGTGGGC 573
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
193 leGlyIysIeuGInGlyValGInArgIleValSerLeuAspAsnGln 209
574 GAAGCAACCATGCTGCTGCCGACGACGCTCCCTCCCTCAGAGGCGG 623
      |||||||:|||||:|||||:|||||:|||||:|||||:
210 GluAlaIThrIleValTyrGInProHISLeuIleSerValGInGluMetly 226
624 GGAAGCGGTATGGTGGATTCTTCGCAACACCTGCTAT 663
      :|||:|||||:|||||:|||||:|||||:|||||:
226 slys.....GInIleGluAlaMetClyPheProAlaPhe 237

seq_name: SwissProt_40:CYF_CHLVU

seq_documentation_block:
ID   CYF_CHLVU          STANDARD:          PRU:          341 AA.
AC   P56316:
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Apocytochrome F precursor.
GN   PETA.
OS   Chlorella vulgaris.
OC   Chloroplast.
OC   Eukaryota; Viridiplantae; Chlorophyta; Trebouxiophyceae; Chlorellales;
OC   Chlorellaceae; Chlorella.
NCBI_TaxID=3077;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=IAM C-27 / TAMIVA.
RX   MEDLINE=97303241; PubMed=9159184;
RA   Wakasugi T., Nagai T., Kapoor M., Sugita M., Ito M., Ito S.,
RA   Tsudzuki J., Nakashima K., Sugita M.,
RA   Inamura A., Yoshinaga K., Sugita M.;
RT   "Complete nucleotide sequence of the chloroplast genome from the
RT   green alga Chlorella vulgaris: the existence of genes possibly
RT   involved in chloroplast division."
RL   Proc. Natl. Acad. Sci. U.S.A. 94:5967-5972(1997).
CC   -1- FUNCTION: TRANSLOCATES PROTONS ACROSS THE THYLAKOID MEMBRANE AND
CC   TRANSFERS ELECTRONS FROM THE RIBSKE IRON-SULFUR PROTEIN AND PASSES
CC   THEM TO PLASTOCYANIN. THIS FUNCTION IS VERY SIMILAR TO THAT OF
CC   MITOCHONDRIAL CYTOCHROME C1.
CC   -1- SUBUNIT: MEMBER OF THE CYTOCHROME B6/F COMPLEX INCLUDING
CC   CYTOCHROME B6, CYTOCHROME F AND PROBABLY AN IRON SULFUR PROTEIN.
CC   -1- SUBCELLULAR LOCATION: Chloroplast thylakoid membrane (Probable).
CC   -1- SIMILARITY: BELONGS TO THE CYTOCHROME C FAMILY.
CC   -----
CC   THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----
DR   EMBL: AB001684; BAA57987.1; -.
DR   HSP: P36438; IHC2.
DR   InterPro: IPR002325; Apocyt_F.
DR   InterPro: IPR000345; CytC_heme_bln.
DR   Pfam: PF01333; Apocytochrome_F; 1.
DR   PRINTS: PR00610; CYTOCHROME_F.
DR   PROSITE: PS00190; CYTOCHROME_C_1.
KW   Electron transport; Heme; Chloroplast; Thylakoid;
KW   Photosynthesis; Photosystem I; Photosystem II; Transit peptide;
KW   Transmembrane.
FT   TRANSIT 1..56 CHLOROPLAST (BY SIMILARITY).
FT   CHAIN 1..341 APOCYTOCHROME F.
FT   METAL 57..57 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
FT   METAL 57..57 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
FT   BINDING 77..77 HEME (COVALENT) (BY SIMILARITY).
FT   BINDING 80..80 HEME (COVALENT) (BY SIMILARITY).
FT   METAL 81..81 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
FT   METAL 81..81 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
FT   TRANSMEM 307..327 POTENTIAL.

```

```

SQ   SEQUENCE 341 AA; 37350 MW; FBC53FC5DE09465F CRC64;
      alignment_scores:
      Quality: 100.50 Length: 312
      Ratio: 0.718 Gaps: 13
      Percent Similarity: 44.872 Percent Identity: 20.513

alignment_block:
US-09-303-518D-569 x CYF_CHLVU ..
Align seg 1/1 to: CYF_CHLVU from: 1 to: 341

7 CGTTTACATTCAGGCTGTTTCCCTTGGCAACGCCATGCATCT 56
  :|||||:|||||:|||||:|||||:|||||:|||||:
6 LysIeuGInPheAsnPheIleProAsnLeuLys.....LysHISAlaVal 20
57 GTTGACCGCCCTGCTCAAAATGCTCTCCCTGCGCTGCTGCTGCG 106
  :|||||:|||||:|||||:|||||:|||||:|||||:
20 LpHeSerPheTrrpGlyGlnAsnGlnAlaSnIleLeuLysPheSerThrLeuV 37
107 ACACGCTGGAAACCGGCTCGACATCTGCGTTTACCTTTAAAGGAA 156
  :|||:|||||:|||||:|||||:|||||:|||||:
37 alSerLysGlyValLeuValLeuValLysSerPhePheThrAlaSer 53
157 GACCGCGCGCGCATCTC.....GCCATATGCGTCA 188
  :|||||:|||||:|||||:|||||:|||||:|||||:
54 SerAsnAlaTyrProIlePheAlaGlnGlnAsnTyrAlaAsnProArgGly 70
189 GCGAGCGATGAT..... 201
  :|||||:|||||:|||||:|||||:|||||:|||||:
70 uAlaAsnGlyArgIleValLysAlaAsnCysHISLeuAlaGluLysProI 87
201 ..... 201
87 leGluIleGluValProGlnAlaValLeuProAspThrValPheGluAla 103
202 .....CCGACCCCAAAACGGCAAGCGCTTTTGGCGAAGC 239
  :|||||:|||||:|||||:|||||:|||||:|||||:
104 ValValLysIleProTyrAspLysGlnIleLysGlnValLeuAlaAsnGly 120
240 GCGCAAGGC.....GGTT 253
  :|||||:|||||:|||||:|||||:|||||:|||||:
120 LysLysGlyAspLeuAsnValGlyValValLeuLeuProAspGlyP 137
254 TGGAACCTGCCCCCGCTTTTTCAGAAACCGGAGACATAGAACAAAG 303
  :|||||:|||||:|||||:|||||:|||||:|||||:
137 heGluIleAlaPro.....ProAspArgIleProGluGlu 148
304 TTCAAGCGGTACACGCTGGCAACATGTGACAGAGCTTTGGACAAACA 353
  :|||||:|||||:|||||:|||||:|||||:|||||:
149 MetLysAlaLysValGlyLysLeuTyrPheGlnProTyrSerAlaGluLys 165
354 CGAAGGCTGCTATTCATCAGCGCGACATCGGACGCTTCGATTGGGCG 403
  :|||||:|||||:|||||:|||||:|||||:|||||:
165 sLysThrIlePheValValGlyProValProGlyLys..... 177
404 GAGCGTACATCAGCAGACAGCTCCGCTCCGCTGACCGCATGTACAA 453
  :|||||:|||||:|||||:|||||:|||||:|||||:
178 ..LysTyr.....SerGluMetValPheProIleLeuSerPro..Asp 190
454 CCGCGCCCAATCAAGCATGACAAATC.....ATGCAAGCGCGG 494
  :|||||:|||||:|||||:|||||:|||||:|||||:
191 ProAlaLysThrLysSerIleSerTyrLeuLysTyrProIleTyrValGly 207
495 CAGGGTGGCGGCAAGAAAGAAACCGGCTTACCACATCAAGGGGTCA 544
  :|||||:|||||:|||||:|||||:|||||:|||||:
207 YGlyAsnArgGlyArgGlyGlnValTyrProAspGlySerLysSerAsn 224
545 AACCAATCAACAAAGCCGCTGCGGACGACCAACATGCTGCGCC 594
  :|||||:|||||:|||||:|||||:|||||:|||||:
224 snThrIlePheThrAlaSerAlaAlaIleLysIleThrAlaIle..... 238

```


AC P54742; 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Serine/threonine protein kinase afsk (EC 2.7.1.-).
 GN AFSK.
 OS Streptomyces griseus.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 CC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
 NCBI_TaxID=1911;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96186909; PubMed=8635757;
 RA Ueda K., Uneyama T., Beppu T., Horinouchi S.;
 RT "The aerial mycelium-defective phenotype of Streptomyces griseus
 resulting from A-factor deficiency is suppressed by a Ser/Thr kinase
 of S. coelicolor A3(2).";
 RL Gene 169:91-95(1996).
 CC -1- FUNCTION: INVOLVED IN THE REGULATION OF SECONDARY METABOLISM BY
 PHOSPHORYLATING, ON BOTH SER AND THR, THE AFSK GLOBAL REGULATORY
 PROTEIN INVOLVED IN THE CONTROL OF SECONDARY METABOLISM (BY
 SIMILARITY).
 CC -1- PTM: AUTOPHOSPHORYLATED AT THR AND SER RESIDUES (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 or send an email to license@isb-sdb.ch).
 CC EMBL: D45246; BAA08203.1;
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR002290; Ser_thr_pkinase.
 DR Pfam: PF00069; pkinase.1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP.1.
 DR PROSITE: PS00108; PROTEIN_KINASE_ST.1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM.1.
 KW Transferase; Serine/threonine-protein kinase; ATP-binding;
 KM phosphorylation.
 FT DOMAIN 16 272 PROTEIN KINASE.
 FT NP_BIND 22 30 ATP (BY SIMILARITY).
 FT BINDING 44 44 ATP (BY SIMILARITY).
 FT ACN_SITE 138 138 BY SIMILARITY.
 SQ SEQUENCE 807 AA; 85231 MW; 66C274219155D091 CRC64;
 alignment_scores:
 Quality: 98.50 Length: 220
 Ratio: 1.026 Gaps: 11
 Percent Similarity: 43.636 Percent Identity: 25.455
 alignment_block:
 US-09-303-518D-569 x AFSK_STRGR ..
 Align seg 1/1 to: AFSK_STRGR from: 1 to: 807
 82 TCCCTGCTGCGCTTCTGTCGACACGCTGGAAACCGGCTC..... 126
 190 SerValThhGlyAlaSerAspIlePheSerLeuGlySerThhLeuValPh 206
 127GACATCTGCGCTTTTACCTTTTAAAGAAAGACCGCGGC 166
 206 ehaIaIaThhGlyHisAlaProPheHis..... 215
 167 GCATGCTGCCAATATGCTGACGACGATATCCCAACCCCAAAAG 216
 216GlyAlaAsnPro..... 219
 217 GTCAAAGCGTTTTCGGAACGCGAAGCGGCTTGGAACTTCCGCC 266
 |||:|||||:|||||: |||

220 ValGluThrValPheMetLeuValArgGluGly.....Pr 231
 267 CGCGTTTTCAGAAACCGGAAGACATAGAAACATGTTCAAGCGGTAC 316
 231 AspLeuGluGlyLeuProAspAspLeuArgProLeuIleGluSerGlysm 248
 317 ACGGCTGGGAACATGTGACGACGCTTTGGACAAACAGAGGCTGCTA 366
 248 etGlnMetAspAlaThrHisProAlaIaGluProArgAspLeuGlnAla 264
 367 TTCATCAGCGCGCACATC.....GGCAGCTACGATTGGCGGACG 407
 265 GlnLeuAlaProHisLeuPheAlaSerGlySerAspSerGly..... 279
 408 CTACATTCAGCCAGCAGCTTCGCTCCGCTGACCGCCATGTACAACGCG 457
 280 ThrAlaSerAlaIaThrPheProValProAlaThrAlaMetIleGluArg 295
 458 CGAAATCAAGACGATAGACAAATATCATGACGCGGCGGCTTCGCGC 507
 296 ArgArgGlyGlyArgArgThr.....Al 303
 508 AAGGAAACCGCGCTACACGATACAGGCGTCAAAACATCATCA 557
 303 ArgArgProProArg..... 308
 558 AGCCCTGCGTTGGGGAAGAACCATGCTGCTGCCGACGCTCCCT 607
 309Pro.ArgProAlaGlyLeuArgAla 317
 608 CCCCTCAAGAGGCGGGAAGCGGTATGGGTGATTTCTTCGCAACCT 657
 317 IapProGlnIaProGlyAlaGlyHisArgLeuAlaGlnArgGlyArgPro 333
 658 GCCTATACCATGACGCTGGCGGCAAAATGGACACGTCGAAAGCGGTGA 707
 334 AlaPheAlaLeuProAlaValLeuAlaIaValaIaArgValArg 350
 708 AACG 711
 350 glnr 351
 seq_name: SwissProt_40:AT7A-CRIGR
 seq_documentation_block:
 ID AT7A-CRIGR STANDARD; PRT; 1476 AA.
 AC P49015;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Copper-transporting ATPase 1 (EC 3.6.3.4) (Copper pump 1)
 GN (Fragment).
 GN ATP7A.
 OS Cricetulus griseus (Chinese hamster).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
 OC Cricetulus.
 NCBI_TaxID=10029;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K1;
 RX MEDLINE=96154683; PubMed=8589689;
 RA Camakaris J., Petris M.J., Bailey L., Shen P., Lockhart P.,
 RT Glover T.W., Barcroft C., Patton J., Mercer J.F.;
 RT "gene amplification of the Menkes (MVK; ATP7A) P-type ATPase gene of
 CHO cells is associated with copper resistance and enhanced copper
 efflux";
 RL Hum. Mol. Genet. 4:2117-2123(1995).
 CC -1- FUNCTION: MAY FUNCTION IN THE EXPORT OF COPPER FROM THE CYTOPLASM
 TO AN INTRACELLULAR ORGANELLE. IT MAY SERVE AS WELL FOR THE EXPORT
 OF OTHER METALS.
 CC -1- CATALYTIC ACTIVITY: ATP + H(2)O + CU(2+)(IN) = ADP + PHOSPHATE +
 CU(2+)(OUT).

CC -1- SUBUNIT: MONOMER.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- TISSUE SPECIFICITY: FOUNDS IN MOST TISSUES EXCEPT LIVER.
 CC -1- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY
 CC (E1-E2 ATPASES).
 CC -1- SIMILARITY: CONTAINS 6 HMA DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U29946; AAB39918.1; -
 DR HSSP: 004656; IAW0.
 DR InterPro: IPR001757; E1-E2-ATPase.
 DR InterPro: IPR001934; HMA.
 DR InterPro: IPR001454; Hydrolase.
 DR Pfam: PF00403; HMA; 6.
 DR Pfam: PF00702; Hydrolase; 1.
 DR PROSITE: PS00154; ATPase_E1-E2; 1.
 DR PROSITE: PS01047; HMA_1; 5.
 DR PROSITE: PS0846; HMA_2; 6.
 KW Hydrolase; Copper transport; Transmembrane; Phosphorylation;
 KW ATP-binding; Metal-binding; Copper; Repeat.
 KW Cytosolic; (POTENTIAL).
 FT DOMAIN 1 642
 FT TRANSMEM 643 665
 FT TRANSMEM 695 717
 FT TRANSMEM 736 760
 FT TRANSMEM 770 788
 FT TRANSMEM 930 952
 FT TRANSMEM 978 998
 FT TRANSMEM 1347 1373
 FT TRANSMEM 1379 1397
 FT DOMAIN 9 75
 FT DOMAIN 172 238
 FT DOMAIN 277 343
 FT DOMAIN 377 443
 FT DOMAIN 479 545
 FT DOMAIN 555 621
 FT MOD_RES 1034 1034
 FT CARBOHYD 674 674
 FT CARBOHYD 685 685
 FT CARBOHYD 887 887
 FT CARBOHYD 953 953
 FT CARBOHYD 1130 1130
 FT CARBOHYD 1134 1134
 FT CARBOHYD 1448 1448
 FT NON_TER 1476 1476
 SO SEQUENCE 1476 AA; 160335 MW; 6B36F5A2AC358CDB CRC64;

alignment_scores:
 Quality: 98.50 Length: 231
 Ratio: 0.794 Gaps: 14
 Percent Similarity: 53.680 Percent Identity: 25.108

alignment_block:

US-09-303-518D-569 x AT7A_CRIGR ..

Align seg 1/1 to: AT7A_CRIGR from: 1 to: 1476

94 CTTTCCTGCTGCACAGCTGGGAACCGGCTCGCA..... 129
 20 IISerGysValArgThrIleGlnGlnLysIleGlnLysGlnGlnGlyI1 36
 130 CATCTGGCGCTTTTAAAGGAAGAACCGGCGCGCATGCTGCCCA 178
 36 eHIShISleLysValSerLeuGlnGlnLysSerAlaThrIleLysGly 53

179 ATATGCGTCAGCGAGCATGAATCCGACCCCAAAAGGTCMAACCGTT 228
 53 sProLysLeu.....GlnThrProLysThrLeuISnLya 65
 229 TTTCGGGAAGCGCAAAAGCGGTTTG.....GAACGTGCCCGCGCTT 272
 66 IleAspAspMetGlyPheAspAlaLeuLysHisAsnAlaAsnProLeu.. 81
 273 TTTCAGAAACCGGAAAGACATGAACAATGTTCAACCGGTACCGGC. 321
 82ProValLeuThrAspThrLeuPheLeuThrAlaAs 95
 322TGGAGACATGTGCAGCAGCGCTTGGCAAAACAGCA 357
 95 erLeuThrLeuProThrAspHisIleGlnSerThrLeuLys..Thrly 111
 358 GGGC.....TGCTATTCATCAGCGCGCATCGGACGCTA 392
 111 sGlyValThrAspIleLysIlePheProGlnLysArgThrLeuAlaValT 128
 393 CGATTGGCGGACGCTACATCAGCAGCAGCTTCGTCGCGTGAACG 442
 128 hIleIleProSerIleValAsnAlaAsnGlnIleLysGlnLeuValPro 144
 443 CCATGTACAA.....ACGCGCAAAATCAAGC..... 470
 145 GluLeuSerLeuGlnThrGlyThrLeuGlnLysArgSerGlyLacysG 161
 471 ..GATAGCAAAATCATCAGCGCGG.....AGGTTTCGCGG 506
 161 IAspHis.SerMetIleGlnAlaGlyGlnValValLeuLysIleLysVa 177
 507 CAAGGAAACCGCGCTACC..... 528
 177 ILeuGlyMetThrCysHisSerCysThrSerThrThrGlnGlyLysIleG 194
 529 ..ACCATACAGAGGTCGAACAAATCATCAAGCGCTTCGCGGGA 576
 194 LysLysGlnGlnGlyValGlnArgIleLysValSerLeuAspAsnGlnGln 210
 577 GCAACCATCGTCGTCGCCGACGACGCTCCCTCCCTCAAGAGCGGGA 626
 211 AlaThrIleValGlyGlnProHisLeuIleSerValGlnGlnIleLysLy 227
 627 AGCGATGAGTGGATTTCTTGGCAAACTGCTAT 663
 227 s.....GlnIleGlnAlaMetGlyPheProAlaPhe 237

seq_name: SwissProt_40:RPOC_CHLTR

seq_documentation_block:
 ID RPOC_CHLTR STANDARD; PRT; 1396 AA.
 AC 084316;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE DNA-directed RNA polymerase beta' chain (EC 2.7.7.6) (Transcriptase
 DE beta' chain) (RNA polymerase beta' subunit).
 GN RPOC OR CT314.
 OS Chlamydia trachomatis.
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_TaxID=813;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=D/UW-3/CX;
 RX MEDLINE=99000809; PubMed=9784136.
 RA Stephens R.S., Kaiman S., Dammel C.J., Fan J., Marathe R., Aravind L.,
 RA Mitchell W.P., Olinger L., Tatusev R.L., Zhao Q., Koonin E.V.,
 RA Davis R.W.;
 RT "Genome sequence of an obligate intracellular pathogen of humans:
 RT Chlamydia trachomatis";
 RL Science 282:754-759(1998).
 CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION

CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
 CC SUBSTRATES.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC (RNA)(N).
 CC -1- SUBUNIT: THE ENZYME CONSISTS OF THE SIGMA CHAIN AND THE CORE
 CC ENZYME WHICH IS COMPOSED OF 2 ALPHA CHAINS, 1 BETA CHAIN, AND 1
 CC BETA' CHAIN.
 CC -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: AE001304; AAC67907.1;
 CC InterPro: IPR000722; RNA_pol_A.
 CC Pfam: PF00623; RNA_pol_A; 1.
 CC Transferrase: DNA-directed RNA polymerase; Transcription;
 CC Complete proteome.
 CC KW
 CC SEQUENCE 1396 AA; 154904 MW; B24BF841D284065 CRC64;

alignment_scores: 98.00 Length: 332
 Ratio: 0.662 Gaps: 15
 Percent Similarity: 44.578 Percent Identity: 20.181

alignment_block:

US-09-303-518D-569 x RPOC_CHLTR ..

Align seg 1/1 to: RPOC_CHLTR from: 1 to: 1396

51 CATCTGTTGACCGCCCTGCTCAATGCGCTCTCCCTGCGCTTCT 100
 |||||
 1084 HSPRO.....GlnllealallethyrAlaAspAlaAsme 1095
 101 GTCTGCACACGCTGGAAACCGGCTCGACACATCTGCGCTTACTTTA 150
 ::::|
 1095 TlysgluLeuValGlyThrTyAlaIleProSeleLYAlaIle.IleSer 1111
 151 AAGGAAGACCGCGCGCATGCTCGCAATATGCGTCCAGCGCANTGAA 200
 |||||
 1112 ValGluGluGlnArgIleAlaProGlyMetLeuAlaArgLeu. 1127
 201 TCCGCAACCCCAACCGTCAAGCCGTTTTCGGAACGCGCAAAAGCG 250
 |||||
 1128ProArg.....GlyAlaIleLysThrLysAspIleThrGlyG 1140
 251 GTTTCG.....GAACCTGCCCCCGCTTTTCAGAAACCGGAA 288
 |||||
 1140 TlyLeuProArgValAlaGluLeuValGluAla.....ArgLysProGlu 1154
 289 GACATAGAAACATG.....TTCAAGCGGT 314
 |||||
 1155 AspAlaIleAspIleAlaLysIleAspGlyValAlaAspPheLysGlyI 1171
 315 ACACGGCTGGGAACATGTCACAGCGCTTGGAC.....A 349
 ::::|
 1171 eGlnLysAsnLysArgIleLeuValAlaArgSerGluIleThrGlyMetG 1188
 350 AACACGAGGCGCTATTTCATCAGCGCCGACATC.....GGC 387
 ::::|
 1188 IugluLunLysLeuIleSerLeuThrLysIleLeuIleValGlnArgGly 1204
 388 AGCTACGATTTGGCGGACGCTACATCAGCCAGACGCTTCCCGGT 437
 ::::|
 1205 AspSerValIleLysGlyGlnGlnLeuThrAspGlyLeuValAlaProH 1221
 438 GACCGCGATG.....TACAAACCGCGGAATCAAGGATAG 475
 ::::|

1221 sgluileugluilecysglyValArgGluLeuGlnLysTyLeuVal 1238
 476 ACAAAATCATCAGCGCGCGAGGTTGGCGC.....AAGGAAA 516
 ::::|
 1238 sngluValGlnGluValTyArgLeuGlnGlyValAspIleAsnAspLys 1254
 517 ACCGCGCTACAGCATCAAGGAGTCACAAATCATCAAGCCCTGCG 566
 ::::|
 1255 HsValGluIleIleValArgGlnMetLeuGlnLysValArgIleThrAs 1271
 567 TTCGCGCAGACCAACCATCTGCTGCGCGACAGCTCCCTCC..... 609
 |||||
 1271 perGlyAspThrThrLeuLeuPheGlyLysPylAspLysGluP 1288
 610CCTCAGAAGCGCGGAGGAGCGGTATGGGTG 639
 ::::|
 1288 heTyGluGluAsnArgArgThrGluGluAspGly..... 1299
 640 GATTTCCTGCGCAACCTGCTATACATGACGCTG..... 675
 |||||
 1300GlyLysProAlaGlnAlaValProValLeuGlyIleTh 1313
 675 675
 1313 rLysAlaSerLeuGlyThrGluSerPheIleSerAlaIleSerPheGln 1330
 676GCGGCAAAATTTGGCAGAC 693
 1330 spThrThrArgValLeuThrAspAlaIleLacysSerSerLysThrAspThr 1346
 694 GTCAAGCGCTGAAAACCTGTTTCTGCTGCAAGCGCTGCGCGG 743
 ::::|
 1347 LeuLeuGlyPheLysGluAsnValIleMetGlyLysMetIleProGlyI 1363
 744 ACNAGGTTTCGATTTCACATCCGCCCTGCCAAGGGAATGAAGCGG 793
 |||||
 1363 yThrGlyPheAspThrHisLysArgIleLysGln.....HisLeuG 1377
 794 ACAAGCCCATGATGCGCGCGCTTTCACACGCAATGCCAATAT 837
 ::::|
 1377 IuLysGluGlnGluAspLeuValPheAspPheAspSerGluPhe 1391
 seq_name: SwissProt_40:ALR_STRCO
 seq_documentation_block:
 ID ALR_STRCO STANDARD: PRT: 391 AA.
 AC 086786;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Alanine racemase (EC 5.1.1.1).
 GN ALR OR SC664.23.
 OS Streptomyces coelicolor.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Saunders D., Harris D., Parkhill J., Barrell B.G., Rastandream M.A.;
 RL Submitted (Aug-1998) to the EMBL/Genbank/DBJ databases.
 CC -1- CATALYTIC ACTIVITY: L-alanine -> D-alanine.
 CC -1- COFACTOR: PYRIDOXAL PHOSPHATE.
 CC -1- SIMILARITY: BELONGS TO THE ALANINE RACEMASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC


```

77 laValGlyGlyProAsnGlnAsnLysAsnPheGlyAsnAsnLysGlyGly 93
252 TTTGGAACTTGGCCCGCGCTTTTCAGAAAACCGGAGACATAGAAAACA 301
94 Phe.....ValGlyAsnArgAsnArgAsnAsn 102
302 TGTTCAAAGCGGTACACG..... 320
102 naaArgAlaGlyAsnGlnAsnArgThrPheProGlyAsnAsnAsnSerA 119
320 ..... 320
119 snGlnLysProAsnAsnGlnThrSerLysAlaAspGlyProAsnAlaLeu 135
321 .....CTGGACATGTGCAGACGCTTTGGACAAACG 355
136 AlAlAsnAsnAsnGlnProAlaThrAlaAlaAlaGlyGlnAsnGlnAlaAs 152
356 AAGGCGTGTATTCATCAGCCGCCACATCGGACGTAAGTTGGGGCGA 405
152 ngLAsnAlaAsnLysGlyLysAsnGlnArgLys.....GlyGlnA 166
406 CGCTACATCAGCCAGACGCTTCCGCTCCGCTGACCGCATGTAACAAC 455
166 snGlnAsnGlnAsnGlnVal.....HisGlyGlnGly 176
456 GCCGAAATTCAAAGCGATACAAATCATGCAGCGCGGAGCGGTTCGG 505
177 AsnGlnGlyGlyProGlyAsnGlnGlyAlaGlyAsnGlnGlyGly 193
506 GCAAGGAAAAACCGCGCTACACATACAAAGGGGTCAA..... 545
193 ngLAsnGlnGlyGlyAlaGlyAsnGlnGlyAsnGlyGlnGlyPheArg 210
546 .....ACAAATCATCAAGCCCTGGTTCGG..... 572
210 LysArgAsnAlaGlyAsnAsnGlnGlyGlyPheSerGlyGlyProGln 226
573 .....CGAAGCAACCATGTCCTGCTCCGACCGACGCTCCCTCCCTC 613
227 AsnGlnGlnArgAsnArgAsnArgSerGlyProArgPro..... 240
614 AAGAAAGGCGGAGCGCTATGGGTGATTTCTTCCGCAACCTGCTAT 663
241 .....GlyGlyGlyAlaGlyAlaMetAsnSerThrAsnMetG 254
664 ACCATGACGCGCGCGCAAAATTTGCGACACGTCAAAGCGTGAACCT 713
254 LysGlyGlyGlyGlyGlyGlyGlyGlyGlyProArgGlyGlyGlnAsp 270
714 GTTTTCTGCTGCGAAGCGCTGCTGCGGAGCAAGTTTCGATTTGCACA 763
271 PheHeLethrGlnArgLysArgSerLysSerGlyProThrPheGln 287
764 TCCGCCCGCTCCAA..... 777
287 euGlnProValGlnValProThrGlnThrLysPheSerGlyLysArgAsnArg 303
778 .....GGGGAATTGAACGCGCAAAAGCCCATGAT 807
304 LeuTyrValGlyAsnLeuThrAsnAspLethrAspAsp 316

seq_name: SwissProt_40:5E5_RAT

seq_documentation_block:
ID 5E5_RAT STANDARD; PRT; 825 AA.
AC 063003:
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE 5E5 antigen.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WISTAR; TISSUE=Brain;
RX MEDLINE=96015159; PubMed=8537300;
RA Suzuki E., Kojima N., Yoshimura K., Uyemura K., Obara K., Akagawa K.;
RT "Cloning and sequence analysis of cDNA for a possible DNA-binding
protein 5E5 in the nervous system.";
RL J. Biochem. 118:122-128(1995)
CC -1- FUNCTION: MIGHT HAVE DNA-BINDING ABILITY.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN NEURONS.
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.ch).
DR EMBL: D37934; BAA07153.1;
KW DNA-binding; Nuclear protein; Antigen.
SQ SEQUENCE 825 AA; 86831 MW; AF667FE2FD55BDF CRC64;

```

```

alignment_scores:
Quality: 97.50 Length: 294
Ratio: 0.780 Gaps: 12
Percent Similarity: 42.517 Percent Identity: 22.449

```

```

alignment_block:
US-09-303-518D-569 x 5E5_RAT ..

```

```

Align seg 1/1 to: 5E5_RAT from: 1 to: 825

```

```

6 TCGTTTCAATTCAGCGCTGTTCCTCCCTTTGCAACCGCATGACATCC 55
535 SerGlnThrLeuProAlaLeuAlaGlyAlaProThrAlaHisAla 551
56 TGTTCAGCGCGCTGTCGAATGCTCTCCCTGCGCTTTCCTGTCG 105
551 aValProGlyProGly.....ProAlaAlaAlaThrLeuGlyG 564
106 CACACGCTGGGAACCGCGCTCGACATCGCGCTTTACCTTTAAAGA 155
564 Ly..... 564
156 AGACCGCGCGCATGTCGCAATGATGCGTCAGGAGCATGATCCG 205
565 .....ArgLysArgLysGlySerThrPargLysGlyArg..... 575
206 ACCCCAAACGTCGAACCGCTTTTGGCAAGCGCAAAAGCGGTTCG 255
576 .....ArgGlyGlyGlyAlaGlyAlaSerGlyGlyGlyArgGly 591
256 GAACCTTCCCGCGCTTTTCAGAAAACCGAAGACATAGAAACAATGT 305
591 Ly.....ArgGlyArgGlyLysArgGlySerGly 601
306 CAAAGCGGTACACG.....CTGGACATGTGCGACGAGCTTTGGACA 349
602 LeuSerGlyThrArgGlnAspAlaGlySerProSerLalaArgArgLysGly 618
350 AACACGAAGCGCTGTATTCATCAGCGCGCAATCGCACGCTACGATT 399
618 uGlnArgArgArgGly..... 623
400 GCGGAGCGTATCATCAGCAGCATTCGCTCCGCTGACCGGCATGTA 449
624 .....HisGlyProProAlaAlaGlyAlaAlaGlnValSerThr 636

```



```

1271 ysglnleuvalaserhislglrlyalhlsmetgluargarglygly 1287
348 CAACAGACGAGGCTGCTATTCATCACGCCGACATCGCAGTACGATT 397
1288 GLythrlyglsalatlhrargluaspargPropharglysglylncy 1304
398 TGGCGGACGCTACATCAGCAGCAGCTTCGTCGGCGGACCGGACG 447
1304 sglYargThrTyr.....ArgHisA 1311
448 TACAACCGCCGAAATCAAGCAGACAAATATCATCGCGCGCAG 497
1311 laqlserleuLeuasnHis.....ArgargserHisglurhrglyln 1325
498 GGTTCGGCGCAAGGAAAAACCGGCTACGACATACAGAGGCTCAAC 547
1326 TysercysProthrhcysProlysthrTyrSerAsnArgMetaLeu 1342
548 AAATCATCAAGCCCTGCGTTCGCGGCAACCATCGTCTCGCCGAC 597
1342 sasrhlslgl.....ArgleuHlsserGluaasnaraydrga 1354
598 CAGCTCCCTCCCTCAAGAGCGGCGGAGCGCTATGGGT..... 638
1354 rghargalaglyargserargargthAlaValArgCysAlaLeucysgly 1370
639 .....GGATTTCCTCGCAACCTGCTACCATGACATGA 670
1371 ArgserPheProglyYarglyserleuGlurghlslleuHlsgly 1387
671 C.....GCTGCGCGCAAAATTCGCACAC 693
1387 uglurhrgluargluProalaaanglylnclyleuaspglythra 1404
694 GTCAAGCGCGTGAACCCGTTTCTGCTCGCGCAACCGCTCGCGGG 743
1404 laAlaserGlualaasn.....LeuThrnglyserGlnclyleu 1417
744 ACAAGTTTCGATTTCGCACATCCG.....CC 769
1418 ThrghlnclylglYalgluProValProHlslleuGluaasglyalpr 1434
770 CCGTCAAGGCGGAAATTGAACGCGGACAGCAAGCCCA 803
1434 oArprogly.....GluaYserglnsrPro 1443

seq_name: SwissProt_40:NTC1_HUMAN

seq_documentation_block:
ID NTC1_HUMAN STANDARD; PRT; 2444 AA.
AC P46531;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Neurogenic locus notch protein homolog 1 precursor (Translocation-
DE associated notch protein TAN-1) (Fragment).
GN NOTCH1 OR TAN1.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91347367; PubMed=181692;
RA Ellisen L.W., Bird J., West D.C., Soreng A.L., Reynolds T.C.,
RA Smith S., Sklar J.;
RT "TAN-1, the human homolog of the Drosophila notch gene, is broken by
RT chromosomal translocations in T lymphoblastic neoplasms.";
RL Cell 66:649-661(1991).
CC -!- FUNCTION: MAY BE IMPORTANT FOR NORMAL LYMPHOCTE FUNCTION. IN
CC ALTERED FORM, MAY CONTRIBUTE TO TRANSFORMATION OR PROGRESSION
CC IN SOME T-CELL NEOPLASMS.

```

```

CC CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC CC -!- TISSUE SPECIFICITY: IN FETAL TISSUES MOST ABUNDANT IN SPLEEN,
CC CC BRAIN STEM AND LUNG. ALSO PRESENT IN MOST ADULT TISSUES WHERE IT
CC CC IS FOUND MAINLY IN LYMPHOID TISSUES.
CC CC -!- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
CC CC -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC CC -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC CC -!- SIMILARITY: CONTAINS 5 ANK REPEATS.
CC CC -----
CC CC this SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
DR DR EMBL: M73980; AA60614.1; -.
DR DR HSPF: P00740; I1XA.
DR DR MIM: 190198; -.
DR DR InterPro: IPR002110; ANK.
DR DR InterPro: IPR000152; Asx_hydroxyl.
DR DR InterPro: IPR000561; EGF-like.
DR DR InterPro: IPR000742; EGF_2.
DR DR InterPro: IPR001881; EGF_Ca.
DR DR InterPro: IPR000800; Notch.
DR DR Pfam: PF00023; ank; 6.
DR DR Pfam: PF00008; EGF; 36.
DR DR Pfam: PF00066; notch; 3.
DR DR SMART: SM00248; ANK; 5.
DR DR SMART: SM00179; EGF_CA; 22.
DR DR SMART: SM00001; EGF_Like; 13.
DR DR SMART: SM00004; NL; 2.
DR DR PROSITE: PS50088; ANK_REPEAT; 4.
DR DR PROSITE: PS50297; ANK_REPEAT_REGION; 1.
DR DR PROSITE: PS00010; ASX_HYDROXYL; 20.
DR DR PROSITE: PS00022; EGF_1; 34.
DR DR PROSITE: PS01186; EGF_2; 26.
DR DR PROSITE: PS01187; EGF_CA; 18.
KW KW Differentiation; Neurogenesis; Repeat; ANK repeat; EGF-like domain;
KW Transmembrane; Signal; Glycoprotein.
KW TM
FT CHAIN 1
FT FT 19 >2444
FT FT 19 1736
FT FT TRANSMEM 1737 1757
FT FT DOMAIN 1758 >2444
FT FT 20 38
FT FT 59 99
FT FT DOMAIN 102 139
FT FT 140 176
FT FT DOMAIN 178 216
FT FT 218 255
FT FT 257 293
FT FT 295 333
FT FT 335 371
FT FT 372 410
FT FT 412 450
FT FT 452 488
FT FT 490 526
FT FT DOMAIN 528 564
FT FT 566 601
FT FT 603 639
FT FT 641 676
FT FT 678 714
FT FT 716 751
FT FT 753 789
FT FT 791 827
FT FT 829 868
FT FT 870 906
FT FT 908 944
FT FT 946 982
FT FT 984 1020
FT FT DOMAIN 1022 1058

```

```

FT DOMAIN 1060 1096 EGF-LIKE 28.
FT DOMAIN 1098 1144 EGF-LIKE 29.
FT DOMAIN 1146 1182 EGF-LIKE 30.
FT DOMAIN 1184 1220 EGF-LIKE 31.
FT DOMAIN 1222 1266 EGF-LIKE 32.
FT DOMAIN 1268 1306 EGF-LIKE 33.
FT DOMAIN 1308 1347 EGF-LIKE 34.
FT DOMAIN 1349 1385 EGF-LIKE 35.
FT DOMAIN 1388 1427 EGF-LIKE 36.
FT REPEAT 1446 1481 LIN/NORCH 1.
FT REPEAT 1482 1523 LIN/NORCH 2.
FT REPEAT 1524 1563 LIN/NORCH 3.
FT REPEAT 1563 1603 ANK 1.
FT REPEAT 1603 1642 ANK 2.
FT REPEAT 1642 1681 ANK 3.
FT REPEAT 1681 1720 ANK 4.
FT REPEAT 1720 1759 ANK 5.
FT DOMAIN 1759 1798 POLY-VAL.
FT DOMAIN 1798 1837 POLY-ARG.
FT DOMAIN 1837 1876 POLY-PRO.
FT DOMAIN 1876 1915 POLY-ALA.
FT DOMAIN 1915 1954 POLY-GLY.
FT DOMAIN 1954 1993 POLY-GLY.
FT DOMAIN 2000 2039 POLY-GLN.
FT DOMAIN 2039 2078 POLY-PRO.
FT DOMAIN 2078 2117 BY SIMILARITY.
FT DOMAIN 2117 2156 BY SIMILARITY.
FT DOMAIN 2156 2195 BY SIMILARITY.
FT DOMAIN 2195 2234 BY SIMILARITY.
FT DOMAIN 2234 2273 BY SIMILARITY.
FT DOMAIN 2273 2312 BY SIMILARITY.
FT DOMAIN 2312 2351 BY SIMILARITY.
FT DOMAIN 2351 2390 BY SIMILARITY.
FT DOMAIN 2390 2429 BY SIMILARITY.
FT DOMAIN 2429 2468 BY SIMILARITY.
FT DOMAIN 2468 2507 BY SIMILARITY.
FT DOMAIN 2507 2546 BY SIMILARITY.
FT DOMAIN 2546 2585 BY SIMILARITY.
FT DOMAIN 2585 2624 BY SIMILARITY.
FT DOMAIN 2624 2663 BY SIMILARITY.
FT DOMAIN 2663 2702 BY SIMILARITY.
FT DOMAIN 2702 2741 BY SIMILARITY.
FT DOMAIN 2741 2780 BY SIMILARITY.
FT DOMAIN 2780 2819 BY SIMILARITY.
FT DOMAIN 2819 2858 BY SIMILARITY.
FT DOMAIN 2858 2897 BY SIMILARITY.
FT DOMAIN 2897 2936 BY SIMILARITY.
FT DOMAIN 2936 2975 BY SIMILARITY.
FT DOMAIN 2975 3014 BY SIMILARITY.
FT DOMAIN 3014 3053 BY SIMILARITY.
FT DOMAIN 3053 3092 BY SIMILARITY.
FT DOMAIN 3092 3131 BY SIMILARITY.
FT DOMAIN 3131 3170 BY SIMILARITY.
FT DOMAIN 3170 3209 BY SIMILARITY.
FT DOMAIN 3209 3248 BY SIMILARITY.
FT DOMAIN 3248 3287 BY SIMILARITY.
FT DOMAIN 3287 3326 BY SIMILARITY.
FT DOMAIN 3326 3365 BY SIMILARITY.
FT DOMAIN 3365 3404 BY SIMILARITY.
FT DOMAIN 3404 3443 BY SIMILARITY.
FT DOMAIN 3443 3482 BY SIMILARITY.
FT DOMAIN 3482 3521 BY SIMILARITY.
FT DOMAIN 3521 3560 BY SIMILARITY.
FT DOMAIN 3560 3599 BY SIMILARITY.
FT DOMAIN 3599 3638 BY SIMILARITY.
FT DOMAIN 3638 3677 BY SIMILARITY.
FT DOMAIN 3677 3716 BY SIMILARITY.
FT DOMAIN 3716 3755 BY SIMILARITY.
FT DOMAIN 3755 3794 BY SIMILARITY.
FT DOMAIN 3794 3833 BY SIMILARITY.
FT DOMAIN 3833 3872 BY SIMILARITY.
FT DOMAIN 3872 3911 BY SIMILARITY.
FT DOMAIN 3911 3950 BY SIMILARITY.
FT DOMAIN 3950 3989 BY SIMILARITY.
FT DOMAIN 3989 4028 BY SIMILARITY.
FT DOMAIN 4028 4067 BY SIMILARITY.
FT DOMAIN 4067 4106 BY SIMILARITY.
FT DOMAIN 4106 4145 BY SIMILARITY.
FT DOMAIN 4145 4184 BY SIMILARITY.
FT DOMAIN 4184 4223 BY SIMILARITY.
FT DOMAIN 4223 4262 BY SIMILARITY.
FT DOMAIN 4262 4301 BY SIMILARITY.
FT DOMAIN 4301 4340 BY SIMILARITY.
FT DOMAIN 4340 4379 BY SIMILARITY.
FT DOMAIN 4379 4418 BY SIMILARITY.
FT DOMAIN 4418 4457 BY SIMILARITY.
FT DOMAIN 4457 4496 BY SIMILARITY.
FT DOMAIN 4496 4535 BY SIMILARITY.
FT DOMAIN 4535 4574 BY SIMILARITY.
FT DOMAIN 4574 4613 BY SIMILARITY.
FT DOMAIN 4613 4652 BY SIMILARITY.
FT DOMAIN 4652 4691 BY SIMILARITY.
FT DOMAIN 4691 4730 BY SIMILARITY.
FT DOMAIN 4730 4769 BY SIMILARITY.
FT DOMAIN 4769 4808 BY SIMILARITY.
FT DOMAIN 4808 4847 BY SIMILARITY.
FT DOMAIN 4847 4886 BY SIMILARITY.
FT DOMAIN 4886 4925 BY SIMILARITY.
FT DOMAIN 4925 4964 BY SIMILARITY.
FT DOMAIN 4964 5003 BY SIMILARITY.
FT DOMAIN 5003 5042 BY SIMILARITY.
FT DOMAIN 5042 5081 BY SIMILARITY.
FT DOMAIN 5081 5120 BY SIMILARITY.
FT DOMAIN 5120 5159 BY SIMILARITY.
FT DOMAIN 5159 5198 BY SIMILARITY.
FT DOMAIN 5198 5237 BY SIMILARITY.
FT DOMAIN 5237 5276 BY SIMILARITY.
FT DOMAIN 5276 5315 BY SIMILARITY.
FT DOMAIN 5315 5354 BY SIMILARITY.
FT DOMAIN 5354 5393 BY SIMILARITY.
FT DOMAIN 5393 5432 BY SIMILARITY.
FT DOMAIN 5432 5471 BY SIMILARITY.
FT DOMAIN 5471 5510 BY SIMILARITY.
FT DOMAIN 5510 5549 BY SIMILARITY.
FT DOMAIN 5549 5588 BY SIMILARITY.
FT DOMAIN 5588 5627 BY SIMILARITY.
FT DOMAIN 5627 5666 BY SIMILARITY.
FT DOMAIN 5666 5705 BY SIMILARITY.
FT DOMAIN 5705 5744 BY SIMILARITY.
FT DOMAIN 5744 5783 BY SIMILARITY.
FT DOMAIN 5783 5822 BY SIMILARITY.
FT DOMAIN 5822 5861 BY SIMILARITY.
FT DOMAIN 5861 5900 BY SIMILARITY.
FT DOMAIN 5900 5939 BY SIMILARITY.
FT DOMAIN 5939 5978 BY SIMILARITY.
FT DOMAIN 5978 6017 BY SIMILARITY.
FT DOMAIN 6017 6056 BY SIMILARITY.
FT DOMAIN 6056 6095 BY SIMILARITY.
FT DOMAIN 6095 6134 BY SIMILARITY.
FT DOMAIN 6134 6173 BY SIMILARITY.
FT DOMAIN 6173 6212 BY SIMILARITY.
FT DOMAIN 6212 6251 BY SIMILARITY.
FT DOMAIN 6251 6290 BY SIMILARITY.

```

```

FT DISULFID 645 655 BY SIMILARITY.
FT DISULFID 650 664 BY SIMILARITY.
FT DISULFID 666 675 BY SIMILARITY.
FT DISULFID 682 693 BY SIMILARITY.
FT DISULFID 697 702 BY SIMILARITY.
FT DISULFID 704 713 BY SIMILARITY.
FT DISULFID 720 730 BY SIMILARITY.
FT DISULFID 725 739 BY SIMILARITY.
FT DISULFID 741 750 BY SIMILARITY.
FT DISULFID 757 768 BY SIMILARITY.
FT DISULFID 772 783 BY SIMILARITY.
FT DISULFID 788 799 BY SIMILARITY.
FT DISULFID 795 806 BY SIMILARITY.
FT DISULFID 800 815 BY SIMILARITY.
FT DISULFID 817 826 BY SIMILARITY.
FT DISULFID 833 844 BY SIMILARITY.
FT DISULFID 838 855 BY SIMILARITY.
FT DISULFID 857 867 BY SIMILARITY.
FT DISULFID 874 885 BY SIMILARITY.
FT DISULFID 879 894 BY SIMILARITY.
FT DISULFID 896 905 BY SIMILARITY.
FT DISULFID 912 923 BY SIMILARITY.
FT DISULFID 917 932 BY SIMILARITY.
FT DISULFID 934 943 BY SIMILARITY.
FT DISULFID 988 999 BY SIMILARITY.
FT DISULFID 993 1008 BY SIMILARITY.
FT DISULFID 1010 1019 BY SIMILARITY.
FT DISULFID 1026 1037 BY SIMILARITY.
FT DISULFID 1031 1046 BY SIMILARITY.
FT DISULFID 1048 1057 BY SIMILARITY.
FT DISULFID 1064 1075 BY SIMILARITY.

```

alignment_scores: quality: 97.00 length: 275
 ratio: 0.815 gaps: 19
 Percent Similarity: 43.273 Percent Identity: 23.636

alignment_block: us-09-303-518d-569 x NTCL_HUMAN ..

Align seg 1/1 to: NTCL_HUMAN from: 1 to: 2444

```

2 TGTGTGTTTACAAATTCAGCGTGTTCCTCCCTTGGCAACGCCA..... 46
||| :||| ||||| |||
89 CysAlaLeuMetLysPheSerGly.....ProLeuGlyThrProLeuAs 103
47 .....TGCACATCCTGTGACCCGCTGCTCAATGCTCTCCCTGC. 88
||| :||| ||||| |||
103 PasaAlaCysLeuThr.....AsnProCysArgAsnGlyGlyThrCysA 118
89 .....TGGCGCTTTCCTGTCGACACACGC 112
118 spleuleuthrLeuthrGlyThrGlyCysArgCysPro.ProGlyTrpse 134
113 TGGGAAA.....CGGCTCGGA 129
:|||||
134 rGlyLysSerCysGlnGlnAlaAspProcysAlaSerAspProcysAlaA 151
130 CATCTGCG...GTTTACCTTTAAAGGAAGACCGCGCGCATCTGCG 176
:||||| ||||| ||||| :|||
151 snGlyGlyGlnCysLeuProPheGlnAlaSerGlyThrCysArgCysPro 167
177 CAATATCGCTCAGGC.....AGGCATGAATCCCGACCCCAAAACGG 217
||| :||| ||||| |||
168 ProSerGlnHisGlyProThrCysArgGlnAspValAsnLeuGlyGly 184
218 TCAGAGCCGTTTTCGCGAAGCGCAAAAGCGGTTTGGAACTTGGCCCC 267
:||||| :|||
184 nLysProArgLeuGlyCysArgGlyGly.....GlyThrCys.... 195
268 GCGTTTTCAGAAACCGAAGACATAGAAACAATGTTCAAGCGGTACA 317
:||||| :|||
196 .....HisAsnGlnValGlySerGlyArg.CysValCysArgAlaThrH 210

```


281 rGAspLeuSerIeuSerGIuAlaIArgIleProValProLysGIuVal... 296
 607 TCCCTCAGAGGCGGAGGATGGGTATGGGTGATTTCTTGGCAACC 656
 297ArgGIuGIuGIuGIuGIuVal... GluArgPheProGIuGIuP 310
 657 T.....GCCATTACCATGACGCTGGCGGCAAAATGG 688
 310 OTYrArGTyrPheAlaIleuTyrArgAlaIleuGIuGIuAlaI 327
 689 CACAGCAAGGCGTGAACCCCTGTTTCTGCTCCGACCCCTGCCCT 738
 327 euserThInGIuGIuGIuAlaIArgAlaIleuValGIuGIuGIu 343
 739 GCGGACAGGTTTC.....GATTTCACATCCGCCCTGCCA 776
 344 GIuGIuValGIuGIuGIuAlaIleuAlaIleuAlaIleuArgProLeuGI 360
 777 AGGGGANTTGAC 789
 360 uAlaIArgLeuSer 364

seq_name: SwissProt_40:H1LR_CHICK

seq_documentation_block:

ID H1LR_CHICK STANDARD; PRT; 218 AA.
 AC P08288;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-AUG-1988 (Rel. 08, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Histone H1.11R.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-87250632; PubMed-3597432;
 RA Coles L.S., Robins A.J., Madley L.K., Wells J.R.E.;
 RT "Characterization of the chicken histone H1 gene complement.
 J. Biol. Chem. 262:9656-9663(1987)."
 CC -1- FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF
 NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M17020; AAA48790.1; -
 DR PIR: C28456; C28456.
 DR HSP: P08287; IGHC.
 DR InterPro: IPR001386; Linker_histone.
 DR Pfam: PF00538; Linker_histone; 1.
 DR SMART: SM00526; H15; 1.
 KW Chromosomal protein; Nuclear protein; DNA-binding; Multigene family.
 FT INIT_MET 0
 FT DOMAIN 37 110 GLOBULAR
 SO SEQUENCE 218 AA; 21672 MW; CB9724BF14654A6 CRC64;

alignment_scores:

Quality: 95.00 Length: 272
 Ratio: 0.709 Gaps: 15
 Percent Similarity: 49.265 Percent Identity: 25.735

alignment_block:

US-09-303-518D-569 x H1LR_CHICK ..

Align seg 1/1 to: H1LR_CHICK from: 1 to: 218

54 CCTGTGACCGCCCTGCTCAAAAGCCTCTCCCTGCTGCTTCTCTGTC 103
 5 ProAlaAlaAlaProAlaAlaAlaProAlaProAlaAla.....Ly 18
 104 TGCACAGCGTGGCAACCGGCGGACATCTGCGGATTCCTTTAAAG 153
 18 sAlaAlaAlaAlaLysProLysAlaAlaLys.....G 30
 154 GAAGACCGCGCGCATCGTCCCAATATGCGTCAGACGATGATCC 203
 30 LyAla.LysAlaArg.....LysProAlaGIuProSerVa 41
 204 CGACCCCAAAACGTCATCAAAACCTTTTGGGAAACGCAAAAGCGGTT 253
 41 LThInGIuLeuIleThrLysAlaValSerLysGIuArgLysGIuL 58
 254 TCGAATCTTGCCCGCGGCTTTTCAGAAACCGAAGACATAGAACATG 303
 58 euserLeuAla.....AlaLeu 63
 304 TTCAAAGCGGTACACGCTGGACATG.....TGCAGCA 338
 64 LysLysAlaLeuAlaAlaGIuGIuGIuGIuGIuGIuGIuGIuGIu 80
 339 GCGTTTGAACAAACAGAGGCGTGTATTCATCAGCGCGGCGCATCGCA 388
 80 rGIuLysLeuGIuLysLysSerLeuValSer.....LysGIu 92
 389 GCTACATTTGGGCGGACGCTACATCAGCCAC...ACCTTCGTTCCG 435
 93 ThrLeuValGIuThrLysGIuThrGIuAlaSerGIuSerPheArgLeu 109
 436 CTGACCGCATGTATCAACCGCGCAAAATCAAGCATAGCAAAATCAT 485
 109 rLysLysProGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIu 121
 486 GCAGCGGCGGAGGTTTCGCGCAAGAAACCGCGCTACGACATAC 535
 122 ..LysLysAlaSerAlaAla.....LysProLysLysAlaAla 134
 536 AAGGCGTCAAAATCATCAAAAGCCTGCGTGGGGAAGCAACCATC 585
 135 Lys.....LysProAlaAlaAlaAlaLysLysProLy 145
 586 GTCTGCGCCGACACGCTCCCTCCCTCAGAGAGCGGGAGCGGATG 635
 145 sLysAlaValAlaValLysLysSerProLysLysAlaLysProAla 162
 636 GGTGATTTCTTGGCAACCTGCTATACCATGACCTGGCGCAAAAT 685
 162 LaSerAlaThrLysSerValLysSerPro..... 172
 686 TGGCAGACGTCAAAGCGCTGAAACCTGTTTCTGCTCGAAGCGCTG 735
 173LysLysAlaAlaLysPro..... 178
 736 CCGCGGCAAGGTTGATTTGCATCCGCGCGGCGGAGGGAATT 785
 179LysLysAlaValAlaAlaLysSerProAlaLysAlaLys 191
 786 GAAGCGGCAAAACCGCATGATGCGCGGTTCACACGCAATGCCGAT 835
 192AlaValLysProLysAlaAlaLysProLysAlaLysProLys 206
 836 ATTGATACGCC 847
 207 AlaAlaLysAla 210

RL Biochemistry 32:11345-11351(1993).
 CC - FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF
 CC NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.
 CC - SUBCELLULAR LOCATION: Nuclear.
 CC - SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M17019; AAA48789.1; -
 DR PIR: B28456; B28456.
 DR PDB: 1GHC; 3I-AUG-94.
 DR InterPro: IPR001386; Linker_histone.
 DR Pfam: PF00538; Linker_histone; 1.
 DR SMART: SM00526; H15; 1.
 KW Chromosomal protein; Nuclear protein; DNA-binding; Multigene family;
 KW Acetylation; 3D-structure.
 FT INIT_MER 0 0
 FT MOD_RES 1 1 ACETYLATION (BY SIMILARITY).
 FT DOMAIN 40 113 GLOBULAR.
 SQ SEQUENCE 224 AA; 22397 MW; D3D057CB97865CAF CRC64.

alignment_scores:
 Quality: 95.00 Length: 267
 Ratio: 0.714 Gaps: 14
 Percent Similarity: 49.813 Percent Identity: 25.843

alignment_block:

US-09-303-518d-569 x H1L1_CHICK

Align seg 1/1 to: H1L1_CHICK from: 1 to: 224

```

78 CCTCTCCGCTGGCGCTTCCGCTGTCACACGCTGGCAAAACCGGCTCG 127
   ||| ||||| ||| ||| ||||| |||||
5 PROIALPROIALAIAIAAIAAIAAIAAIAAIAAIAAIAAIAAIAAIAA 21
128 GACATCTGGCGTTTACCTTTAA.....GGAAGACGGCGGCGC 168
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
21 salaiaiaiaiaiaiaiaiaiaiaiaiaiaiaiaiaiaiaiaiaia 37
169 ATCGTGGCAATATGCGTCAGCAGCATGAATCCGACCCCAAAACGCT 218
   :: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
38 .....LysProAlaGlyProSerValThGluLeuIleTh 49
219 CAAGCGCTTTTGGGAAACGCAAAAGCGGTTTGAACCTTGGCCCGC 268
   :||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
49 rlysaiaiaiaiaiaiaiaiaiaiaiaiaiaiaiaiaiaiaiaiaia 64
269 CTTTTTTCAGAAACCGGAAGACATGAACAATGTTCAAGCGGTACAC 318
65 .....AlaLeuLysLysAlaLeuAl 71
319 GGCTGGGAACATG.....TCGACGAGGCTTTGGACAAACA 353
   :||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
71 aiaiaiaiaiaiaiaiaiaiaiaiaiaiaiaiaiaiaiaiaiaia 88
354 CGAAGCGCTATTCATCAGCGCCCAATGCGAGCTACATTTGGGCG 403
   ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
88 eulysSerLeuValSer.....LysGlyThrLeuValGlnThr 100
404 GACGCTACATCAGCCAGC...AGCTTCGCTCCGCTGACCGCATGTAC 450
   ::||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
101 LysGlyThrGlyAlaSerGlySerPheArgLeuSerLysProGly 117
451 AAACGCGCAAAATCAAGCATAGCAAAATCATGACGCGGCGAGGCT 500
   ::||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
117 uValysLysLysAlaProLys.....LysLysAlaSerA 129

```

seq_name: SwissProt_40:VE2_HPV05

seq_documentation_block:

ID VE2_HPV05 STANDARD; PRT; 514 AA.

```

AC P06921;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Regulatory protein E2.
GN E2.
OS Human papillomavirus type 5.
OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
OC NCBI_TaxID=10578;
OX [1]
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87207670; PubMed=3033892;
RA Zachow K.R., Ostrow R.S., Farias A.J.;
RT "Nucleotide sequence and genome organization of human papillomavirus
RT type 5."
RL Virology 158:251-254(1987).
CC -1- FUNCTION: E2 REGULATES VIRAL TRANSCRIPTION AND DNA REPLICATION.
CC IT BINDS TO THE E2RE RESPONSE ELEMENT (5'-ACCNANNNGGT-3') PRESENT
CC IN MULTIPLE COPIES IN THE REGULATORY REGION. IT CAN EITHER
CC ACTIVATE OR REPRESS TRANSCRIPTION DEPENDING OF E2RE'S POSITION
CC WITH REGARDS TO PROXIMAL PROMOTER ELEMENTS. REPRESSION OCCURS
CC BY STERICALLY HINDERING THE ASSEMBLY OF THE TRANSCRIPTION
CC INITIATION COMPLEX. THE E1-E2 COMPLEX BINDS TO THE ORIGIN OF DNA
CC REPLICATION.
CC -1- SUBUNIT: BINDS DNA AS A DIMER.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M17463; AAA46986.1; -
DR PIR: D26277; W2W15.
DR HSP: P03122; 2BOP.
DR InterPro: IPR001866; E2_N.

```

DR InterPro: IPR000427; Early2.C.
 DR Pfam: PF00511; E2.C; 1.
 DR Pfam: PF00508; E2.N; 1.
 DR ProDom: PD000672; Early2.C; 1.
 DR ProDom: PD000678; E2.N; 1.
 DR Early protein; transcription regulation; Activator; DNA-binding;
 KW Trans-acting factor; DNA replication; Repressor; Nuclear protein.
 SEQUENCE 514 AA; 57532 MW; 0100722143BF7256 CRC64;

alignment_scores:

Quality: 95.00 Length: 197
 Ratio: 0.888 Gaps: 8
 Percent Similarity: 54.315 Percent Identity: 24.365

alignment_block:

US-09-303-518D-569 x VE2_HPVO5 ..

Align seg 1/1 to: VE2_HPVO5 from: 1 to: 514

```

156 AGACCGCGCGGCATCGTCCCAATATGCGTCAGGAGCATGATCCG 205
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
267 ArgSerArgSerArgSerArgSerArgSerArgSerArgSerArg 283
    . . . . .
206 ACCCAACGCGTCAAGCGTTTGGGAAACGCGCAAGCGGCTTG 255
    . . . . .
283 SerThrThrThrThrThrThrThrThrThrThrThrThrThr 298
    . . . . .
256 GAACCTGCCCCCGCTTTTTCAGAAACCGAAGACATAGAACATGTT 305
    . . . . .
298 ThrSerThrThrThrThrThrThrThrThrThrThrThrThr 314
    . . . . .
306 CAACGCGTACGCGCTGGGAAACATGCGACAGCTTGGCAACACG 355
    . . . . .
315 ArgSerProThr.....ThrGSAArgArgGlyGlyArgSerP 328
    . . . . .
356 AAGGCTGCTATTCATCAAGCGCGCATCGGACGAGTATGCGGGA 405
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
328 Arg.....ArgArgSerArgSerProSer 337
    . . . . .
406 CCCTCATCAGCAGCAGCTTCCGTCGACCGGACCATGTAACAC 455
    . . . . .
337 hr.....SerSerSerCysThrThrGlnArgSerGlnArg 349
    . . . . .
456 GCGGAATATCAAGCGATAGCAAAATCATGCGAGCGGAGCTTCG 505
    . . . . .
350 ArgAlaGluSerSerThrThrArgGlyAlaArgGlySerArgGly 366
    . . . . .
506 GCAAGAGAAACCGCGCTACCATCAAGAGGCTCA..ACAATC 552
    . . . . .
366 GGLyGlySerArgGlyArgGlyArgGlyArgGlyArgSerSer 383
    . . . . .
553 ATCAAGCGCTGCGTTCGCGGCAAGCAACCATCGTCCCGGACG 602
    . . . . .
383 SerSerProAlaIleLysArgSerArgGlySerAlaLysLeu 399
    . . . . .
603 CCCCTCCCTCAAGAGCGGGAAGCGCTATGAGTGTCTTCGCGA 652
    . . . . .
400 GlyValSerPro.....GlyGlyValGlySerLeuArg 412
    . . . . .
653 AACCTGCTATACCATGAGCGTGGCGCAAAATGCG.....ACAG 696
    . . . . .
412 ValSerSerLysThr.....GlyArgLeuGlyArgLeuLeuGlu 427
    . . . . .
697 AAGCGGTGAACCGCTTTTCTGCGAAGCGCGCTCC 737
    . . . . .
427 IuAlaArgSerProValIleIleValIleValIleValIle 440
    . . . . .

```

seq_name: SwissProt_40:LIPE_AERHY

seq_documentation_block:

ID LIPE_AERHY STANDARD; PRT; 684 AA.
 AC P40600;

DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Extracellular lipase precursor (EC 3.1.1.3) (Triacylglycerol lipase).
 GN Lip.
 OS Aeromonas hydrophila.
 CC Bacteria; Proteobacteria; gamma subdivision; Aeromonadaceae;
 CC Aeromonas.
 CX NCBI_Taxid=644;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN=H3;
 RX MEDLINE=93378423; PubMed=8368830;
 RA Angula J., Rodriguez Aparicio L.B., Naharro G.;
 RA "Purification, gene cloning, amino acid sequence analysis, and
 RT expression of an extracellular lipase from an Aeromonas hydrophila
 RT human isolate.";
 RT Appl. Environ. Microbiol. 59:2411-2417(1993).

CC -1- FUNCTION: THE OPTIMUM CHAIN LENGTHS FOR THE ACYL MOTIETY IS C6 FOR
 CC ESTER HYDROLYSIS AND C6 AND C8 FOR TRIACYLGLYCEROL HYDROLYSIS. THE
 CC OPTIMAL TEMPERATURE FOR ENZYME ACTIVITY IS 37 DEGREES CELSIUS.
 CC HOWEVER, AFTER 30 MIN. AT 55 DEGREES CELSIUS, THE ENZYME RETAINS
 CC 90% OF THE LIPASE ACTIVITY.
 CC -1- CATALYTIC ACTIVITY: Triacylglycerol + H(2)O = diacylglycerol + a
 CC fatty acid anion.
 CC -1- SUBUNIT: MONOMER.
 CC -1- SUBCELLULAR LOCATION: Secreted.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: S65123; AAB28083.2; -;
 DR InterPro: IPR000734; Lipase.
 DR PROSITE: PS00120; LIPASE_SER; 1.
 KW Hydrolyase; Lipid degradation; Signal.
 FT SIGNAL 1 48
 FT CHAIN 49 684 POTENTIAL.
 FT ACT SITE 568 568 EXTRACELLULAR LIPASE.
 FT ACT SITE 568 568 CHARGE RELAY SYSTEM (BY SIMILARITY).
 SQ SEQUENCE 684 AA; 71903 MW; 325418D3936FA428 CRC64;

alignment_scores:

Quality: 95.00 Length: 319
 Ratio: 0.736 Gaps: 20
 Percent Similarity: 40.439 Percent Identity: 25.705

alignment_block:

US-09-303-518D-569 x LIPE_AERHY ..

Align seg 1/1 to: LIPE_AERHY from: 1 to: 684

```

42 CGCATGACATGCT...GTGACCGCGCTGCTCAATGCGCT..... 80
    ||||| ..||| ||||| ||||| ..|||
272 ArgHisLeuThrProGlyArgAspArgProThrCysLeuProAsp 288
    . . . . .
81 .....CTCCCTGCTGCGCTTCTCTGTCGACACGCT.... 113
    . . . . .
288 aGlyCysArgAlaValProAlaAlaGluGlnLysAspAlaIleAla 305
    . . . . .
114 .....GGGAACCGCGCTCGACATCTGGCGTTTAAAGGAGA 158
    . . . . .
305 IapheGlyProValLeuArgSerThrArgLeuLys.....Arg 318
    . . . . .
159 CCGCGCGCG.....CATCG 172
    |||||
319 ProArgSerIleProValProSerSerCysLeuThrSerProHis 335
    . . . . .
173 TCGCCATATGCGTACGAGCAGCATGATCCGACCCCAAAAGGCTCAA 222
    . . . . .

```



```

272 TTTTCAGAAACCGGAGACATAGAAACATGTTCAAAACGGCTACAGCC 321
      :::::::::: :::::::::: :::::::::: ::
1892 ...LysArgSerProLysHisArgSerLysSerArgLysArgLysArgL 1907
      :: :: :: :: :: :: :: :: ::
322 TGGACACATGTCGACAGGCTTTGGACAAACAGCAGGCTGCTATTCAT 371
      :: :: :: :: :: :: :: :: ::
1907 ysaArgSerSerArgAspAsnArgLysThrValArgAlaArgSerArg 1923
      :::::::::: :::::::::: :::::::::: ::
372 CAGCCGACATGGCAGCTACGATTTGGCGGAGCGCTACATACGCCAGC 421
      :::::::::: :::::::::: :::::::::: ::
1924 ThrProSerArgSerArgSerHisThrProSerArgArgArgSer 1940
      :::::::::: :::::::::: :::::::::: ::
422 AGCTTCGCTTCCGCTGACCGCCATGTACAAACCGCGGAAATCAAGCG 471
      :: :: :: :: :: :: :: :: ::
1940 rArgSerValGlyArgArgArgSerPheSerIleSerProSerArgArg 1957
      :: :: :: :: :: :: :: :: ::
472 ATGACAAATTCATGACGGCGGAGGCTCCGCGCAAGGAAACCGCG 521
      :: :: :: :: :: :: :: :: ::
1957 eraArg.....ThrProSerArgArgSerArg 1965
      :::::::::: :::::::::: :::::::::: ::
522 GCCTACACGATACAAAGGGGTCAAAACATCATCAAGCCCTGCTCGG 571
      :: :: :: :: :: :: :: :: ::
1966 ThrProSerArgArgSerArgThrProSerArgArgSerArgThrPro 1982
      :: :: :: :: :: :: :: :: ::
572 GCGAAGCAACATCGTCTGCTGCGGACCGCTCCCTCCCTCAAGAGCG 621
      :: :: :: :: :: :: :: :: ::
1982 rArgArgSerArgThrProSerArgArgSerArgThrProSerArgArg 1999
      :: :: :: :: :: :: :: :: ::
622 GCGGAAGCGCTATGGGTGATTTCTTCGGAACCGCTATACCATGAC 671
      :: :: :: :: :: :: :: :: ::
1999 rArgArgSerArg..... 2002
      :::::::::: :::::::::: :::::::::: ::
672 GCTGCGGCGCAAAATTTGGCACACGTCAAGCGCTGAACCGCTGTTTCT 721
      :: :: :: :: :: :: :: :: ::
2003 .....SerValValArgArgArgSerPheSerIleSerProValArg 2017
      :: :: :: :: :: :: :: :: ::
722 G.....CTGCGAAGCGCTGCGCTGCGGACAGAGTTTCATTTGCCATC 765
      :: :: :: :: :: :: :: :: ::
2017 rArgArgSerArgThrPro.....LeuArgArgArgPheSerArg 2031
      :: :: :: :: :: :: :: :: ::
766 CGGCC.....CGTCCAGAGGGAATTAACGGCGGCAACAGCCCA 803
      :: :: :: :: :: :: :: :: ::
2031 eProlLeuArgArgLysArgSerArgSerSerGluArgLysArgSerPro 2047
      :: :: :: :: :: :: :: :: ::

```

seq_name: SwissProt_40:SGS3_DROER

seq_documentation_block:

ID SGS3_DROER STANDARD; PRT; 328 AA.

AC P13730;

DT 01-JAN-1990 (Rel. 13, Created)

DT 01-AUG-1991 (Rel. 19, Last sequence update)

DT 01-OCT-1996 (Rel. 34, Last annotation update)

DE Salivary glue protein Sgs-3 precursor.

OS SGS3.

OC Drosophila erecta (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

NCBI_TaxID=7220;

PN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=8332966; PubMed=3138416;

RA Martin C.H., Mayeda C.A., Meyerowitz E.M.;

RT "Evolution and expression of the Sgs-3 glue gene of Drosophila.";

J. Mol. Biol. 201:273-287(1988)

RL -1- DEVELOPMENTAL STAGE: PRODUCED BY THIRD-INSTAR LARVAE.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation-

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: M14652; -; NOT_ANNOTATED_CDS.

DR PIR: S01359; S01359.

DR FLYBASE: FBgn0012266; Dere\Sgs3.

KW Repeat: Signal.

FT SIGNAL 1 23

FT CHAIN 24 328 SALIVARY GLUE PROTEIN SGS-3.

SQ SEQUENCE 328 AA; 36355 MW; 62F27F18C0D8272 CRC64;

alignment_scores:

Quality: 94.50 Length: 231

Ratio: 0.836 Gaps: 14

Percent Similarity: 48.918 Percent Identity: 27.706

alignment_block:

US-09-303-518D-569 x SGS3_DROER ..

Align seg 1/1 to: SGS3_DROER from: 1 to: 328

```

2 TGTTCGTTTACATTCAGCGCTGTTCCCGCT.....TGCGA 39
  ||| :: :: :: ||| ::::::::::
37 CysThrThrAlaArgProThrCysAlaProValThrThrThrCysAr 53
  :: :: :: :: :: :: :: :: ::
40 ACCGCCATGCACATC...TGTGACGCGCTGCTCAATGCTCTCCCT 86
  :: :: :: :: :: :: :: :: ::
53 gProThrThrThrThrArgCysProProProThrThrThrArgCysProp 70
  :: :: :: :: :: :: :: :: ::
87 GCTGCGCTTCTCTGTC...TGACACGCTGGAACCGGCTCGACATC 133
  :: :: :: :: :: :: :: :: ::
70 rOrProThrArgProAlaGluCysThrAla...ThrThrLysArgProThr 85
  :: :: :: :: :: :: :: :: ::
134 TGGCGTTTACCTTTTAAAGAGACCGCGCGCATGTCGCCAATATG 183
  ||| ||| ||| :: :: :: :: :: :: :: ::
86 AlaArgProThrThrThrArgArgThrThrValArgAlaThrThr..... 99
  :: :: :: :: :: :: :: :: ::
184 GGTGAGGAGCATGAATCCCGACCCCAAAAGCGTCAAGCGTTTTCG 233
  :: :: :: :: :: :: :: :: ::
100 ...LysArgAla.....ThrThrArgArgThrThrLysArgAla 112
  :: :: :: :: :: :: :: :: ::
234 GAAACGCGCAAAAGCGGTTTGAACCTTGCCCGCGCTTTTTCGAAAC 263
  :: :: :: :: :: :: :: :: ::
112 hrThrArgArgThrThrVal..... 118
  :: :: :: :: :: :: :: :: ::
284 CGGAAGACATGAAACATGTTCAAGCGGTACAGCGCTGGCAATGTG 333
  ||| ||| ||| :: :: :: :: :: :: :: ::
119 ArgAlaThrThrThrLysArgAlaThrThrArgArgThrThrThrLysArgAl 135
  :: :: :: :: :: :: :: :: ::
334 CAGCAGCGCTTGGACAAACAGAGGCGCTGCTATTC.....T 371
  :: :: :: :: :: :: :: :: ::
135 aProThrArgArgThrThrThrLysArgAlaThrThrArgArgAsnPro 152
  :: :: :: :: :: :: :: :: ::
372 CAGCCGACATCG.....GCAGCTACGATTTGGCGGAGCGCTAC 412
  :: :: :: :: :: :: :: :: ::
152 hrArgArgThrThrThrArgArgAlaProThrLysArgAlaThrThrLys 168
  :: :: :: :: :: :: :: :: ::
413 TCAGCCGACAGCTTCGTTCCCGGTGACCGCCCATGTACAAACCGCGGAA 462
  :: :: :: :: :: :: :: :: ::
169 ArgAlaThrThrThrArgAsnProThrLysArgLysThrThrThrArgArg 185
  :: :: :: :: :: :: :: :: ::
463 ATCAAAACGATAGCAAAATCATGACGGCGGAGGTTGCGGCAAGG 512
  :: :: :: :: :: :: :: :: ::
185 rThrValArgAlaThrThrLysThrThrLysArgAla.....ThrThrLys 200
  :: :: :: :: :: :: :: :: ::
513 AAAAACCG.....CGCTACACGATACAGGG 541
  :: :: :: :: :: :: :: :: ::
200 rGlaProThrLysArgAlaThrThrLysArgAlaProThrLysArgVal 216
  :: :: :: :: :: :: :: :: ::
542 TCAAAACATCATCAAGCGCTGCGTTGGGG.....AAGCAACG 582
  :: :: :: :: :: :: :: :: ::
217 ThrThrLysArgAla...ProThrLysArgAlaThrThrLysArgAlaPr 232

```

583 ATGCTCTGCGCCGACGACGTC...CCTCCCTCAAGAGCG 622
 |||||
 232 CTHTLysArgAlaThrLysArgAlaProThrLysArgAla 246

seq_name: SwissProt_40:SQHC_ZYMMO

seq_documentation_block:
 ID SQHC_ZYMMO STANDARD; PRT; 658 AA.
 AC P3390;
 DT 01-FEB-1994 (rel. 28, Created)
 DT 01-FEB-1994 (rel. 28, Last sequence update)
 DT 16-OCT-2001 (rel. 40, Last annotation update)
 DE Squalene--hopene cyclase (EC 5.4.99.-).
 GN SHC.
 OS Zymomonas mobilis.
 OC Bacteria; Proteobacteria; alpha subdivision; Sphingomonadaceae;
 OC Zymomonas.
 OX NCBI_TaxID=542;
 RX MEDLINE=95202074; PubMed=7894707;
 RA Reipen I.G., Poralla K., Sahm H., Sprenger G.A.;
 RT "Zymomonas mobilis squalene-hopene cyclase gene (shc): cloning, DNA
 sequence analysis, and expression in Escherichia coli."; Microbiology 141:155-161(1995).
 RL [2]
 RN SEQUENCE FROM N.A.
 RP STRAIN-ATCC 31821 / ZM4 / CP4;
 RA Reipen I.G., Sahm H., Sprenger G.A.;
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 31821 / ZM4 / CP4;
 RA Um H.W.;
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: CATALYZES THE CYCLIZATION OF SQUALENE INTO HOPENE.
 CC -1- PATHWAY: KEY ENZYME IN HOPANOID (TRITERPENOID) METABOLISM.
 CC -1- SUBCELLULAR LOCATION: Membrane-associated (by similarity).
 CC -1- SIMILARITY: BELONGS TO THE TERPENE CYCLASE/MUTASE FAMILY.
 CC -1- SIMILARITY: CONTAINS 4 PPTB REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X80766; CAAS6749.1; -
 DR EMBL: X73561; CAAS1958.1; -
 DR EMBL: AJ001401; CA04735.1; -
 DR EMBL: AF203881; AAF12829.1; -
 DR PIR: S37494; S37494.
 DR HSSP: P33247; 15QC.
 DR InterPro: IPR001330; Prenyltrans.
 DR InterPro: IPR002365; Terpene_synth.
 DR Pfam: PF00432; prenyltrans; 4
 DR PROSITE: PS01074; TERPENE_SYNTHASES; 1.
 KW Isomerase; Membrane; Repeat.
 FT REPEAT 69 110 PPTB 1.
 FT REPEAT 418 459 PPTB 2.
 FT REPEAT 486 526 PPTB 3.
 FT REPEAT 534 584 PPTB 4.
 SQ SEQUENCE 658 AA; 74068 MW; 26B6AC3C1BD625C CRC64;

alignment_scores:
 Quality: 94.50 Length: 206
 Ratio: 1.050 Gaps: 9
 Percent Similarity: 43.689 Percent Identity: 22.330

alignment_block:
 US-09-303-518D-569 x SQHC_ZYMMO ..
 Align seg 1/1 to: SQHC_ZYMMO from: 1 to: 658

181 ATGCTCTGCGCCGACGACGTC...CCCAAAAGGTCAAGCCGT 227
 :::::::::: :: ::::::::::
 106 LeuLysMetIleGlyAspSerProAspAlaProHisMetLeuArgAlaAr 122
 228 TTTTGCGGAAAGCGCAAAAGCGGTTTGGAACCTGCCCGCGCTTTTCA 277
 ::::::::::
 122 gAnGluIleuAlaArgGlyAlaMetArgAlaAsnValPheThrA 139
 278 GAAACCGGAAGACATAGAAACATGTTCAAAAGCGGTACAGCGGTGGAA 327
 ::::::::::
 139 rg.....IleGlnLeuAlaLeuPheGlyAlaMetSerTrpGlu 151
 328 CATGTGCGACAGCGTTTGACAAACAGAGGCGTGTATTCATCAGCGC 377
 ::::::::::
 152 HisValProGlnMetProValGlu.....LeuMetLeuMetPr 164
 378 G.....CACATGCGAGCTACGATTTGGCGGAGCGCTACA 412
 164 oGluTrpPheProValHisIleAsnLysMet.....AlaTrpT 177
 413 TCAGCCAGCAGCTTCGCTCCGCTGACCGCCATGACAAACGCCGAA 462
 ::::::::::
 177 rPalaArgThrValLeuValProLeuLeuVal..... 187
 463 ATCAAGCGATGACAAATCATGACGGCGGCGGCTCGGGCAAGG 512
 187 187
 513 AAAACCGCGCTACACGATACAAAGGGTCAACAAATCATCAAGCC 562
 ::::::::::
 188LeuGlnAlaLeuLysProValAlaArgAsnA 198
 563 TGGCTTCGGGCGAAGCA...ACCATGCTCTGCCGACACAGCTCCCGTC 609
 ::::::::::
 198 rgArGlyIleLeuValAspGluLeuPheValProAspValLeuProThr 214
 610 CTCACGAAGCGGGGAGGCGGTATGGTGGATTTCTTGGGC..... 651
 ::::::::::
 215 LeuGlnLysSerLysProIleTrpArgArgPhePheSerAlaLeuAs 231
 652AACTGCTATPACCATGACGCTGGCGGCA 682
 231 pLysValLeuHisLysValGluProLysProLysAsnMetArgAlaL 248
 683 AATTGGCAGACGTCAAAGGGGTAAACCTGTTTCTGTCGCAAGCC 732
 248 ysaLalleHis.....SerCysValHisPheValThrGluArg 260
 733 CTGCTGCGCGGACAAAGT 750
 261 LeuAsnGlyGluAspGly 266

seq_name: SwissProt_40:YT44_STRFR
 seq_documentation_block:
 ID YT44_STRFR STANDARD; PRT; 395 AA.
 AC P20188;
 DT 01-FEB-1991 (rel. 17, Created)
 DT 01-FEB-1991 (rel. 17, Last sequence update)
 DT 16-OCT-2001 (rel. 40, Last annotation update)
 DE Hypothetical 44.4 kDa protein in transposon TN4556.
 OS Streptomyces fradiae.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.
 OX NCBI_TaxID=1906;
 RN [1]
 RP SEQUENCE FROM N.A.

RC TRANPOSON-Tn4556;
 RX MEDLINE=90185236; PubMed=2155856;
 RA Siemieniak D.R., Slightom J.L., Chung S.T.;
 RT "Nucleotide sequence of Streptomyces fradiae transposable element
 Tn4556: a class-II transposon related to Tn3.";
 RL Gene 86:1-9(1990).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M29297; AAA8562.1; -;
 DR PIR; J00430; J00430.
 KW Hypothetical protein; Transposable element.
 SQ SEQUENCE 395 AA; 44379 MW; C50B556F0E3B2838 CRC64;

alignment_scores:

Quality: 94.00 Length: 294
 Ratio: 0.783 Gaps: 13
 Percent Similarity: 40.816 Percent Identity: 22.109

alignment_block:

US-09-303-518d-569 x YT44_STRFR ..

Align seg 1/1 to: YT44_STRFR from: 1 to: 395

```

51 CATCTGTTGACGGCCCTGCTCAAAATGCTCTCCCTGCGCGCTTTCCT 100
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
51 HISHIALASPGLNGLNLEUAGLARGALASPGLYALGLNPROVA 67
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
101 GTCTGCACAGCTGGGAAACCGGCTGACATCTGCGCTTTTACCTTTA 150
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
67 IASPLNHSGLY.....ProProcySLNGLYGLY..... 78
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
151 MAGGAAGACCGCGCGCATCTGCGCAATATGCTGACGAGCATGAA 200
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
79 .....ArgValGlnArgArgGlyGlnGlnGlnGlnArgGlyAsp 91
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
201 TCCGACCCCAAAAGCGTCAAGCGCTTTTCCGAAAGCGCAAAAGCG 250
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
92 ArgArgGlyGlnLeuValProGlnHISHisLeuLeuAspGlyArg 106
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
251 GTTGGAACTGCCCCCGCTTTTTCAGAAAACCGAAGACATAGAACA 300
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
107 .....ProHisProPhGlnAlaLeuHisPheArgArgArg 119
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
301 ATGTTCAAGCGGTACAGCGCTGGGAACATGTGCAGCAGGCTTT 344
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
119 SPLN.....ProGlyValGlyValAspLeuLeuAlaArgGlyValHis 134
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
345 .....GACAAACAGCAAGGCGTCTATTCATCACGCCCG 379
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
135 ASPLeuGlnArgGlyArgGlnLeuArgArgProLeuHisAlaArg 151
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
380 ACATGCGGACCTAGCATTTGGCGGACCTACATCAGCAGAGCTTCCG 429
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
151 nArgLeuGlnLeuLeuArgAlaLeuAlaLeuArgProGlyGlnGln 168
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
430 TTTCCGTCAGCGCCATGTACAAACCGCGCA.....AATCAAGC 470
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
168 eArgArgArgProGlnHisLeuGlnGlnValGlnArgValAlaHisArgPro 184
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
471 GATAGCAAAATCATGACAGCGGCGGCTTCCGCGCAAGAAAGAAACCG 520
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
185 ArgArgGlnValHisGlyGlyGlnGlnArgArgGln..... 197
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
521 CGCTACCAAGATACAGGCGTCAAAACAAATCATCAAGCCCTGCTCG 570
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

```

```

198 .....A 198
571 GCGGAGCAACATCTCTCTCCCGACACATCTCCCTC..... 608
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
198 lAlaSerValGlyValProAlaGlnGlnArgArgLeuAlaHisPro 214
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
609 .....CCCTCAGACGCGGGAAGCGGTATGGG 637
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
215 ValProGlyGlnLeuArgGlnProAlaArgArgTyrProArgHisLeu 231
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
638 TGGATTCTTCGGAACCTGCTTACCATGACCTGCGGCGCAAAATG 687
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
231 nGly.....AlaHisSerGlnArgGlyAspLeuVal 242
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
688 GCACACGTCAAGCGGTGAACCTGTTTCTGCTGCAACGCGCTCC 737
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
242 rArgProLeuGlnArg.....ProLeuHisLeu..... 250
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
738 TCGCGGACAGGTTTGCATTCACATCGCCCGCCGCAAGGGATGGA 787
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
251 .....ArgThrArgHisProHisArgProAlaProGlnProValGln 264
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
788 ACGGCGCAAAACCCATGATGCCCGCTGTTCACACG..... 824
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
264 nLeuArg.....AspProLeuProGlyArgHisArgProAlaGlySer 280
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
825 .....CAATGCCGAATATGATACCGCG 848
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
280 eThrAlaArgCysProValLeuSerIlePro 290
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

```

seq_name: SwissProt_40:GLT3_WHEAT

seq_documentation_block:

```

ID GLT3_WHEAT STANDARD; PRT; 660 AA.
AC P08488;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Glutenin, high molecular weight subunit 12 precursor.
OS Triticum aestivum (Wheat).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;
OC Triticeae; Triticum.
OX NCBI_TaxID=4565;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. CHINESE SPRING;
RX MEDLINE=86041862; PubMed=3840588;
RA Thompson R.D., Bartels D., Harberd N.P.;
RT "Nucleotide sequence of a gene from chromosome 1D of wheat encoding a
RL HMW-glutenin subunit."
CC Nucleic Acids Res. 13:6833-6846(1985).
CC -!- FUNCTION: GLUTENINS ARE THE HIGH MOLECULAR WEIGHT SEED STORAGE
CC PROTEIN OF WHEAT ENDOSPERM. THOUGHT TO BE RESPONSIBLE FOR THE
CC VISO-ELASTIC PROPERTY OF WHEAT DOUGH.
CC -!- SUBUNIT: DISULFIDE-BRIDGE LINKED AGGREGATES.
CC -!- MISCELLANEOUS: GLUTENINS ARE CODED BY SEVERAL GENES ON EACH OF THE
CC GROUP 1 CHROMOSOMES OF WHEAT.
CC -!- MISCELLANEOUS: THE MATURE PROTEIN IS CHARACTERIZED BY A LARGE
CC NUMBER OF WELL PRESERVED REPEATS OF THE TWO MOTIFS: GQDPGQ AND
CC GQDPGQGGQYTPS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X03041; CA26847.1; -;
DR PIR; A24266; A24266.
DR HSP; P01088; 1BFA.

```



```

456 GCCGAATATCAAGCATAGACAAATATCATGACGGCGGAGGTTGCG 505
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
198 .....GlnGlnSerSerGlyArgGlyHisGlnGlyAlaHisGlnGlu 212
506 GCAAGG.....AAAAACCGCGCTACCAACATA 534
      ::::::::::: ::::::::::: :::::::::::
535 CAAGGGGTCAACAAATCAT..... 554
      ::::::::::: ::::::::::: :::::::::::
229 eiaRghisGlnAlaSPSerProArgValSerAlaArgSerGly 245
555 .....CAAGCCCT...GGTTGGGGGAGCAACCATGCTCTGC 592
      ::::::::::: ::::::::::: :::::::::::
246 GlyArgGlyGlnSerProAlaSerGlyArgSerSerAlaArgAs 262
593 CGGACGACGT...CCGCTCCCTCAGAGA 617
      ::::::::::: :::::::::::
262 pArgProArgGlnProSerProSerGln 271

```

seq_name: SwissProt_40:YHL1_EBV

seq_documentation_block:

```

ID YHL1_EBV STANDARD: PRT; 660 AA.
AC P03181.
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Hypothetical BHLP1 protein.
OS Epstein-Barr virus (strain B95-8) (human herpesvirus 4).
OC Viruses; dsDNA viruses; no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Lymphocryptovirus.
OX NCBI_TaxID=10377;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=84270667; PubMed=6087149;
RA Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,
RA Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,
RA Tufnell P.S., Watell B.G.;
RT "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";
RL Nature 310:207-211(1984).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: V01555; -, NOT_ANNOTATED_CDS.
DR PIR: A03742; Q0BE3.
KW Hypothetical protein; Early protein; Repeat.
FT DOMAIN 149 648 4 x 125 AA RANDOM REPEATS.
FT REPEAT 149 273 1.
FT REPEAT 274 398 2.
FT REPEAT 399 523 3.
FT REPEAT 524 648 4.
SQ SEQUENCE 660 AA; 66244 MW; 86DA1D67A37152A2 CRC64;

```

alignment_scores:

```

Quality: 93.00 Length: 282
Ratio: 0.823 Gaps: 11
Percent Similarity: 40.071 Percent Identity: 25.177

```

alignment_block:

US-09-303-518D-569 x YHL1_EBV

Align seg 1/1 to: YHL1_EBV from: 1 to: 660

27 TCCCCCTTTGGAAACGGCATGACATCTGTTGACGCGCTGCT..... 71

```

211 ThrProHisProGluArgGlySerGlyProAlaAspProProAlaAla 227
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
72 .....CAATGCGCTCTCCCTGTCGCG 93
      ::::::::::: :::::::::::
227 aArgLeuProProGluArgGlnGluProArgLeuProGlnsPleuAla 244
94 CTTTCTGCTGTCGACACCGCTGGAAACCGGCTCGACATCTGGCGTTTAA 143
      ::::::::::: ::::::::::: :::::::::::
244 AlaAlaGlnArgCysProAlaGlyProProThrArgSerGlyAla... 259
144 CTTTAAAGAAAGACCGCGCGCATCTGCGCCATATGGCTCAGGAG 193
260 .....AlaAlaGlnArgThrHisArgArgProProGlyCysProAr 273
194 GCATGAATCCGACCCCAAAACGGTCAAGCGTTTGGCGAAACGGCA 243
273 gSerAlaArgAsnPro..... 278
244 AAGCGGTTTGGAACTTGGCCCCCGCTTTTTCAGAAAACCGAAGACAT 293
279 .....GlyCysProArg.....ThrTrpArgArg 286
294 AGAAACANTGTTCAA..... 308
287 ArgSerGlyAlaGlnArgGlyHisProProProGlyAlaGlyAlaArgPr 303
309 .ACGCTACACGCGGTGGGACATGTGCAGCAGCGTTTGACAAACAGAA 357
      ::::::::::: :::::::::::
303 oSerGlyProThrGlyArgProAlaAlaProGlyAlaProGlyThr 320
358 GGGTGGCATTCATCAGCCGCGACATGGCAGCTACATTTGGCGGACG 407
320 roAlaAlaProGlyProGlyGlyAlaAlaAlaProSerGlyAlaThr 336
408 CTACATCAGCCAGCAGCTCCGTCCTCCGCTGA..CCGCGATGTCAAAACG 456
337 ProHisProGluArgGlySerGlyProAlaAspProProAlaAlaAla 353
457 CCGAAATCAAAAGCATAGACAAATCATGACGCGGCGAGGTTGCGCG 506
353 gLeuProProGluArg..... 358
507 CAAGGAAACCGCGCTTACGACATACAAAGGGTCAAAACATCATCA 556
359 .....GlnGluProArgLeuPro..... 364
557 AAGCCTGCGTGGCGGAGCAACATCGTCTGCGCCGACACGTCGCC 606
365 GlnAspLeuAlaAlaGlnArg.....CysProAlaGlyProPr 378
607 TCCCCCTAAGAGCGGGGAGGCGTAT.....G 635
378 oProThrArgSerGlyAlaAlaAlaGlnArgThrHisArgArgProG 395
636 GGTGATTTCTTGGCAACCTGCTCT...ATACCATGACGCTGGCGGAA 682
      ::::::::::: :::::::::::
395 LysProArgSerAlaArgAsnProGlyCysProArgThrTrpArgArg 411
683 AATTGGCACACGTCAAGCGCGTAAACCCCTGTTTCTGTCGCAACGC 732
      ::::::::::: :::::::::::
412 Arg.SerGlyAlaGlnArgGlyHisProProProGlyAlaGlyAlaArg 428
733 CTGCTGGCGGACAGTTTCGATTTGCGACATCCGCGCGCTGCA 776
      ::::::::::: :::::::::::
428 roSerGlyProThrGlyGlyArgProAlaAlaProGlyAlaPro 442

```

seq_name: SwissProt_40:POLN_RRV1

seq_documentation_block:

```

ID POLN_RRV1 STANDARD: PRT; 1148 AA.
AC P13888.
DT 01-JAN-1990 (Rel. 13, Created)

```

DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Nonstructural polyprotein [contains: Nonstructural protein NSP3 and NSP4] (Fragment)
 OS Ross river virus (strain T48) (RRV)
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
 OC Alphavirus.
 NCBI_TaxID=11032;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88206074; PubMed=2834873;
 RA Straus E.G., Levinson R., Rice C.M., Dalrymple J., Straus J.H.;
 RT "Nonstructural proteins nsP3 and nsP4 of Ross River and O'Nyong-Nyong
 RT viruses: sequence and comparison with those of other alphaviruses."
 RL Virology 164:265-274(1988).
 RN [2]
 RP SEQUENCE OF 1066-1148 FROM N.A.
 RX MEDLINE=83039346; PubMed=6291034;
 RA Ou J.-H., Rice C.M., Dalgaro L., Straus E.G., Straus J.H.;
 RT "Sequence studies of several alphavirus genomic RNAs in the region
 RT containing the start of the subgenomic RNA."
 RL Proc. Natl. Acad. Sci. U.S.A. 79:5235-5239(1982).
 CC -1- PPM: SPECIFIC ENZYMAIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
 CC -1- MISCELLANEOUS: READTHROUGH OF THE TERMINATOR CODON UGA OCCURS
 CC BETWEEN THE CODONS FOR 531-PHE AND 532-LEU.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL: M20539; AAA47407.1; ALT_SEQ.
 DR EMBL: K00046; AAA47403.1; ALT_SEQ.
 DR PIR: A28614; A28614.
 DR InterPro: IPR002589; A1PP.
 DR InterPro: IPR001788; RNA_dep_RNAPol2.
 DR Pfam: PF01661; A1PP; 1.
 DR Pfam: PF00978; RNA_dep_RNAPol2; 1.
 DR SMART: SM00506; A1PP; 1.
 DR Polyprotein; Nonstructural protein.
 FT NON_TER 1 1
 FT CHAIN 1 537 NONSTRUCTURAL PROTEIN NSP3.
 FT CHAIN 538 1148 NONSTRUCTURAL PROTEIN NSP4.
 FT NON_TER 1148 1148
 SQ SEQUENCE 1148 AA; 127154 MW; BD372FF29FDF8FF CRC64;

alignment_scores:
 Quality: 93.00 Length: 272
 Ratio: 0.769 Gaps: 15
 Percent Similarity: 44.485 Percent Identity: 21.691

alignment_block:

US-09-303-518D-569 x POLN_RRV

Align seg 1/1 to: POLN_RRV from: 1 to: 1148

30 CCCTTGGACGCGCATGACATCTCTGTGACCGCCGCTCA..... 74
 420 Procythserlutyralysprollepolaalprarglaalalalr 436
 75ATGCTCTCCCTGCTGCGCTTCCGTCTGCACACGCTGGGAAC 120
 436 gvalvalpvalprola..... 442
 121 CGGCTCGACATCTGGGTTTACCTTTTAAAGAAAGACGCGCGCAT 170
 443Prodylllegln 446
 171 CGTGGCAATATGCTGACGACGATGAATCCGACCAACCAACGCTCA 220

447 Arg.AlaSerThrTyralValSerProThrProThrProThrProValLeuA 463
 221 AAGC...GTTTGGCGAAGACGCAAAAGCGGTTTGAACCTGCCCC 267
 463 rglaserValcysSerValThrThrSerAlaGlyValcyluProThr 479
 268 GCGTTTTCAGAAACCGAGACATAGAAACATGTTCAAGCGGTACA 317
 480 Ala.....ProGluSpluGluValleuThrGluProValH 492
 318 C.....CGTGGCA..... 327
 492 scylsMetArgLuproValGluLeuProThrProGluSpluValA 509
 328CATGTCGACAGCT... 342
 509 splleGlnPheGlyAspPheGluThrSerAspLyslleGlnPheGlyAsp 525
 343 TTGGACAAACGACGAGGCTGCTATTCATCAGCCGACATCGGACGTA 392
 526 IleAspPheAspGlnPheLeuGlyArgAlaGlyAlaTyrIlePheSerSe 542
 393 CGATTGGCGGACGCTACATCAGCCAG.....CAGTTCCGT 430
 542 rAspThrGlyProGlyHisLeuGlnLysSerValArgGlnHisAla 559
 431 TCCCGTGACCGCCATGTAC.....AAACGCGG 459
 559 euProCysGluMetLeuTyrValHisGluGluGluArgThrTyProPro 575
 460 AAAATC...AAAGCATAGACAAATATCATCAGCGGCGGAGGTGGCG 506
 576 AlaLeuAspGluAlaArgGluLysLeuGlnAla..... 587
 507 CAAGGAAACGCGCGCTAC..... 528
 588 LysMetGlnMetAlaProThrGluAlaAsnLysSerArgTyGlnSerA 604
 529 .ACCATACAAAGGGGTCAACAA...ATCATCAAGCCCTGCGTGGCG 573
 604 rGlyValGluAsnMetLysAlaValIleIleAspArgLeuLysAspGly 620
 574 GAAGCAACCATCTGCTGCCCGACAGCTCCCTCCCTCAACAAAGCGG 623
 621 AlaArgThrTyLeuThrGluGlnSerGluLysIlePro..... 633
 624 GAAGCGGTATGGGTGATTTCTGCGCAACCTGCTATACATGACGC 673
 634ThrTyValSerLysTyProArgProValTySerProSer 648
 674 TGCGGCAAAATG 687
 648 alGluAspSerLeu 652

seq_name: Swissprot_40:POLN_RRVN

seq_documentation_block:

ID POLN_RRVN STANDARD; PRT; 2479 AA.
 AC P13887;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Nonstructural polyprotein [contains: Nonstructural proteins NSP1 TO NSP4].
 OS Ross river virus (strain NB5092) (RRV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
 OC Alphavirus.
 NCBI_TaxID=11031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88179556; PubMed=2833022;
 RA Faragher S.G., Meek A.D.J., Rice C.M., Dalgaro L.;

alignment_scores:

Quality: 92.50 Length: 281
 Ratio: 0.690 Gaps: 13
 Percent Similarity: 47.687 Percent Identity: 20.641

alignment_block:

US-09-303-518D-569 x VP10_RDVF ..

Align seg 1/1 to: VP10_RDVF from: 1 to: 353

```

211 AAACGGTAA...GCCGTTTGGGAAACGCAAAAGCGGTTTGA 257
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
32 LysLysValLysLeuGlyThrLeuAlaAsnGlnSerGlyAlaAsnAs 48
258 ACTGCGCCCGCGTTTTCAGAAACCGCAACATAGAAACAATGTTCA 307
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
48 nlethrgluAlaPhePheAspLysLeuArgAsnGlnArgLysSerG 65
308 AAGCGGTACAGCGCTGGGAACATGTGACACAGGCTTG...GACAAA 351
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
65 luAlaThrLeuAlaSerAspLeuAlaGluArgGluLeuThrArgSp 81
352 CAGCAAGGCGCTGCTATCATCATCGCCGCGCATCGGCTACGATTGG 401
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
82 HisLysAlaIleValPheValThrLys.....SerValLeuLeuG 95
402 CGAGCGCTCATCATGACGACAGCTTCGCTCCGCTGACCGCC..... 444
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
95 yglLysSerLeuLysAspLeuLeuProtyrGlyValIleValGlyAla 112
445 ..ATGACAAACCGCGCAAAATCAAGCAGATACAAATCATGACGCG 492
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
112 heilePheIleProGluThrAlaSerValLeuAspAsnVal..... 125
493 GCGAGGTTCCGCGCAAGAAACCGCGCTACGACGATCAAGGGGT 542
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
126 ..ArgValMetIleGlyAsnGlnLysArgProLeuThrValAlaLeu 141
543 CAACAAATCATCAAAAGCGCTGCGTTCG..... 570
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
141 elystrMetAlaLysSerLeuAsnCysAspLeuValGlyAspSerTyr 158
571 .....GGCAAGCAACCATC... 585
158 spThrPheTyrTyrCysAsnSerSerAlaTyrGlyLysAsnLeuLe 174
586 GTCTGCGCCGACCAAGCTCCCTCCCTCAGAAAGCGGGAAGCGGTG 635
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
175 ValSerGluAsnAspPheSerAsnProGlnArgAlaLeuLeuSerVal 190
636 GGTGATTTCTTCGCAAACTGCTATACATGACGCTGCGGCAAAAT 685
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
191 .....GlyAspLeuCysTyrGln.....AlaAlaArgS 200
686 TGGCAGCGTCAAAAGCGTGAACCCGTTTTCGTGGGAACGCGCTG 735
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
200 erlleHisValAlaAlaAlaAsnTyrIleArgIle...PheAspArgMet 215
736 CCGTGGCGAGCAAGGTTTGCATTCGACATCCGCCCGTCAAGGGAAT 785
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
216 ProProGlyPheGlnProSerLysHisLeuPheArgIleIleGlyVal 232
786 GAAC..... 789
232 uasPheGluThrLeuLysThrMetValThrSerAsnIleAlaArgLup 249
790 .....GGCGCAAAAGCGCATGATGCGCGC 813
249 roGlyMetPheSerHisAspAsnValLysAspValLeuHisArgThrGly 265
814 GTGTTCACCGCAATGCCGAATAT.....TGGATACGCGC 848
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

266 ValPheSerProAsnHisHisPheSerAlaValIleLeuTpr...ArgG 281
 849 TTTTCCGACGAGTATCTGTTATGTACACCGCTACAAATG 891
 281 yThrPheSerThrTyrAlaTyrMetPheAsnGlnGluGlnLeu 295

seq_name: SwissProt_40:SPR6_HUMAN

```

seq_documentation_block:
ID   SPR6_HUMAN          STANDARD;          PRT;          344 AA.
AC   Q13247; Q13244; Q13245; Q9UJB8;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Splicing factor, arginine/serine-rich 6 (Pre-mRNA splicing factor
DE   SRP55).
GN   SPR56 OR SRP55.
OS   Homo sapiens (human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX   NCBI_TaxId=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC   TISSUE=Colon;
RX   MEDLINE=96016206; PubMed=7556075;
RA   Sreelton G.R., Caceres J.F., Mayeda A., Bell M.V., Plebanski M.,
RA   Jackson D.G., Bell J.I., Krainer A.R.;
RT   Identification and characterization of three members of the human SR
RT   family of pre-mRNA splicing factors."
RL   EMBO J. 14:4336-4349(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
RA   Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,
RA   Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
RA   Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.D.,
RA   Buck D., Burdall W., Butler A.P., Carder C., Carter N.P.,
RA   Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
RA   Clegg S., Cobley V.E., Collier R.E., Connor R., Corby N.R.,
RA   Coulson A.G., Coville G.J., Deadman R., Dhami P., Dunn M.,
RA   Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
RA   Grahame D.V., Griffiths C., Griffiths M.N.D., Gilliam R., Hall R.E.,
RA   Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
RA   Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
RA   Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
RA   Levaslaiho M.H., Leverhwa M., Lloyd C., Lloyd D.M., Lovell J.D.,
RA   Marsh V.L., Martin S.L., McComachie L.J., McKay K., McMurtry A.A.,
RA   Oliver K., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
RA   Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
RA   Phillimore B.J.C.T., Prichardlingam S.R., Plumb R.W., Ramsay H.,
RA   Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Shownkeen R., Sims S.,
RA   Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
RA   Swann M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
RA   Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
RA   Whitehead S.L., Whitlaker P., Willey D.L., Williams L., Williams S.A.,
RA   Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
RA   Rogers J.;
RT   The DNA sequence and comparative analysis of human chromosome 20."
RL   Nature 414:865-871(2001).
RN   [3]
RP   SEQUENCE OF 21-27 AND 47-55.
RX   MEDLINE=92249775; PubMed=1577277;
RA   Zahler A.M., Lane W.S., Stolk J.A., Roth M.B.;
RT   "SR proteins: a conserved family of pre-mRNA splicing factors."
RL   Genes Dev. 6:837-847(1992).
RN   [4]
RP   FUNCTION: PLAYS A ROLE IN CONSTITUTIVE SPLICING AND CAN MODULATE
RP   THE SELECTION OF ALTERNATIVE SPLICING SITES.
RN   [5]
RP   SUBCELLULAR LOCATION: Nuclear.
RN   [6]
RP   ALTERNATIVE PRODUCTS: 3 ISOFORMS; SRP55-1 (SHOWN HERE), SRP55-
RP   2 AND SRP55-3; ARE PRODUCED BY ALTERNATIVE SPLICING AND DIFFER IN
RP   THEIR C-TERMINAL DOMAINS.
RN   [7]
RP   PTM: EXTENSIVELY PHOSPHORYLATED ON SERINE RESIDUES IN THE RS
RP   DOMAIN (BY SIMILARITY).
CC   CC

```

CC -1- SIMILARITY: CONTAINS 2 RNA RECOGNITION MOTIFS (RRM).
 CC -1- SIMILARITY: BELONGS TO THE SR FAMILY OF SPLICING FACTORS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@isb-sdb.ch).
 CC
 CC EMBL: U30883; AAA93073.1; -
 CC EMBL: U30828; AAA93071.1; -
 CC EMBL: U30829; AAA93072.1; -
 CC EMBL: AL031681; CAB43960.1; -
 CC MIM: 601944; -
 CC InterPro: IPR000504; RRM.
 CC Pfam: PF00076; rrm; 2.
 CC SMART: SM00360; RRM; 2.
 CC PROSITE: PS00030; RRM_RNP_1; FALSE_NEG.
 CC Nuclear protein; RNA-binding; mRNA splicing; Alternative splicing;
 CC Repeat; Phosphorylation.
 CC
 CC DOMAIN 1 72 RNA-BINDING (RRM) 1.
 CC DOMAIN 87 90 GLY-RICH (HINGE REGION).
 CC DOMAIN 110 183 RNA-BINDING (RRM) 2.
 CC DOMAIN 184 343 ARG/SER-RICH (RS DOMAIN).
 CC VARSPLIC 86 135 SGGGGYSSRRTSGRDYGPPREYRLIYENLSSRCMODL
 CC KDFRQAGE -> MINGAEVSTSEAKMTAPPPMPLFHTLC
 CC DCPMTLMLTLPEAMTTAAFC (IN ISOFORM SRP55-
 CC 2).
 CC MISSING (IN ISOFORM SRP55-2).
 CC RSYPPPKATSRSRSRSSRSRSPD -> LKLG
 CC VARSPLIC 313 344 RFSOGTESLYSLASC (IN ISOFORM SRP55-3).
 CC CONFLICT 64 64 H -> R (IN REF. 2).
 CC SEQUENCE 344 AA; 39567 MW; 74AD1CD4C18A7A74 CRC64;
 CC
 CC alignment_scores:
 CC Quality: 92.00 Length: 228
 CC Ratio: 0.948 Gaps: 9
 CC Percent Similarity: 42.544 Percent Identity: 21.491
 CC
 CC alignment_block:
 CC US-09-303-518d-569 x SFR6_HUMAN ..
 CC
 CC Align seg 1/1 to: SFR6_HUMAN from: 1 to: 344
 CC
 CC 105 GCACACGCTGGGAACCGGCTCGACATCTGCGTTTACCTTTAAAG 154
 CC ||| ||| ||| :|||:|||||:|||||:|||||:
 CC 139 ALaARpAlAnIeLysGluAtgThrAsnGluValIleGluPheArgSe 155
 CC 155 A..... 155
 CC 155 rTytSerArMeTtLysAtgAlaLeuAspLysLeuAspLysThrGluIleA 172
 CC 156 AGACCGCGC...GCGCATCGT 173
 CC 172 snGluYrGsnIleArgLeuIleGluAspLysProArgThrSerHisArg 188
 CC 174 CGCGAATATGCGTCAGACGCGATGATCCGACCCCAAAAGGTCMAAG 223
 CC |||:|||||:|||||:|||||:|||||:
 CC 189 ArgSerTyrSerGlySerArgSerArgSerArg..... 199
 CC 224 CGTTTTCGGAACGCGCAAAAGCGGTTTGAACTTTCGCCCGCGTTT 273
 CC :|||:|||||:|||||:|||||:
 CC 200SerArgArgArgSerArgSerArgSerArg..... 209
 CC 274 TTCAGAAAACCGGACATAGAAACATGTTCAACGCGTACCGGCTG 323
 CC :|||:|||||:|||||:|||||:
 CC 210ArgSerArgSerArgSerArgSerArgSerArgSer 224
 CC 324 GGAACATGTGCACAGGCTTTGGACAAACAGCAGGCGCTGATTCATCA 373

225 ArgSerArgSerArgSerArgSerArgSerArgSerArgSerArgSer..... 238
 CC :|||:|||||:|||||:|||||:
 CC 374 CGCGACATCGGACGCTACGATTTGGGGGAGCGGTACATGACGACAG 423
 CC |||||:|||||:|||||:|||||:
 CC 239LysGluYrGsnIleArgSerArgSerArgSerArgSer 248
 CC 424 CTTCGCTCCGCTGACCGCATGTACAAACCGCGAAATCAAGCAT 473
 CC ||| ||| :|||:|||||:|||||:|||||:
 CC 248 ySerLysPro.....LysSerAsp 254
 CC
 CC 474 AGCAAAATCATGACGCGGCGGCTTCGCGCAAGAAACCGCGC 523
 CC |||:|||||:|||||:|||||:|||||:
 CC 255 ArgGlySerHisSer.....HisSerArgSerArgSerArgSerAsp1 268
 CC |||||:|||||:|||||:|||||:|||||:
 CC 524 CTACGACATACAAAGGCGTCAAAATCATCAAGCCCT..... 563
 CC 268 uTyGluLysSerArgSerArgSerArgSerArgSerProLysGluAsnG 285
 CC 564GCGTTCGGGCGAAGCAACCATCGTCCT 590
 CC 285 LysGluLysPheLysSerArgSerArgSerArgSerArgSerArgSer 301
 CC 591 GCCCGACACGTCCTCCCTCC..... 611
 CC ||| |||:|||||:|||||:|||||:
 CC 302 AsnSerProLeuProValProProSerLysAlaArgSerValSerProPr 318
 CC 612TCAAGAGCGCGGGAAG 629
 CC |||||:|||||:|||||:|||||:
 CC 318 oPolysArgAlaThrSerArgSerArgSerArg 329
 CC
 CC seq_name: SwissProt_40:DBPA_RAT
 CC
 CC seq_documentation_block:
 CC ID DBPA_RAT STANDARD; PRT; 361 AA.
 CC AC 062764; 063748; -
 CC DT 15-JUL-1999 (Rel. 38, Created)
 CC DT 15-JUL-1999 (Rel. 38, Last sequence update)
 CC DT 30-MAY-2000 (Rel. 39, Last annotation update)
 CC DE DNA-binding protein A (Cold shock domain protein A) (Muscle Y-box
 CC protein YB2) (Y-box binding protein-A) (Ryb-A).
 CC GN CSDA OR DBPA OR YB2.
 CC OS Rattus norvegicus (Rat).
 CC OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
 CC OX NCBI_TaxID=10116;
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RC STRAIN=SPRAGUE-DAWLEY;
 CC RA Goldman D., Gao J., Burmeister M., Sappu M.;
 CC RT "Characterization of muscle Y-box proteins that bind the mACHR
 CC delta subunit promoter.";
 CC RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
 CC RN [2]
 CC RP SEQUENCE FROM N.A.
 CC RC STRAIN=Wistar; TISSUE=Liver;
 CC RX MEDLINE=94301785; PubMed=8029009;
 CC RA Ito K., Tsutsumi K., Kuzumaki T., Gomez P.F., Otsu K., Ishikawa K.;
 CC RT "A novel growth-inducible gene that encodes a protein with a
 CC conserved cold-shock domain.";
 CC RL Nucleic Acids Res. 22:2036-2041(1994).
 CC CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC CC -1- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; 1 (SHOWN HERE) AND 2;
 CC CC ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC CC -1- TISSUE SPECIFICITY: ABUNDANT IN THE SKELETAL MUSCLE, SPLEEN, AND
 CC CC FETAL LIVER.
 CC CC -1- SIMILARITY: BELONGS TO THE COLD-SHOCK DOMAIN (CSD) FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

```

CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U27893; AAB60520.1; -.
DR EMBL: D28537; BAA05907.1; -.
DR HSSP: P15277; IMC.
DR InterPro: IPR002059; Cold_shock.
DR Pfam: PF00313; CSD; 1.
DR PRINTS: PRO0050; COLDSHOCK.
DR ProDom: PD000621; Cold_shock; 1.
DR SMART: SM00357; CSP; 1.
DR PROSITE: PS00352; COLD_SHOCK; 1.
KW Transcription regulation; Repressor; DNA-binding; Nuclear protein;
KW Alternative splicing.
FT DOMAIN 85 149 CSD.
FT VARSPIC 184 252 MISSING (IN ISOFORM 2).
FT FT 14 14 L -> H (IN REF. 2).
FT FT 52 74 SPGGDADPGAPAPASAPASDEA -> APAPASPPAPGLI
FT FT 52 74 SPRKRKR (IN REF. 2).
FT FT 120 121 PR -> HV (IN REF. 2).
SQ CONFLICT 361 AA; 38651 MM; C6799D5A3DA5C3F3 CRC64;

```

alignment_scores:	Quality: 92.00	Length: 256
	Ratio: 0.821	Gaps: 14
Percent Similarity:	44.800	Percent Identity: 24.4000
alignment block:		
us-09-303-518D-569 x DBPA_RAT		..

Align seg 1/1 to: DBPA_RAT from: 1 to: 361

```

103 GCGACACGCTGGGGAACCGCGCGTCCGACACATCTGGCGTTTACTCTTTAAAG 154
104 GCGACACGCTGGGGAACCGCGCGTCCGACACATCTGGCGTTTACTCTTTAAAG 154
144 AAGtunlaalaalaasnvalThnGlyProAspGlyVal...ProValGluG 159
155 A...AGACCGCGCGCGCATGTGCGCAATTCGCTCA.....GGCA 192
159 ySerhIrrTyrrAlaAlaAsprIargIargIrrTyrrArgGlyTyrrTyGly 176
193 GGCATGAATCCCGACCCCAAAAC.....GGTCAAGCGCGTTTGGCGGA 226
176 rgarIargGlyrProProIargAsnTyrrAlaGlyIugIugIugIugIugLy 192
237 AACGGCAAAAGCGGTTTGGAACTTGGCCCCCGCGGTTTTCAGAAACCGG 286
193 SerGlySerSerGluGlyPheGluProProAlaAla..... 204
287 AAGACATAGAAACAATCTTCAAAAGCGGTACACGGCTGGGGAACATGTGCAG 336
205 .....AspGlyGlnPheSerIrrTyrrAlaArg..... 212
337 CAGGCTTTGGACCAACACGGAAGGCGTGCATTCATCATGCCCGCACATCGG 386
213 .....AsnGlnLeuAlaIrrArgProGln...TyrIrrProIrrTyrrArg 225
387 CAGCTACGATTT.....GGCGGAGCGCTATCATGCAGAC 421
226 GlnIrrArgIrrPheProIrrTyrrHisValGlyGlnIrrPheAspIrrArgSe 242
422 AGGTTCCGTTCCCGCGTACGCCGCATGTACAAACGCCGGAAMAATCAAGCG 471
242 IrrTyrrAlaPheProIrrHisProIrrAsnArgMetGlnIrrAlaGlyIrrLeuGlyIrr 259
472 ATTAGCAAAATATCATGCAGCGCGGCGAGGTTGCCGGCAAGAAAACCGC 521
259 eLlysAspGlyValrProGluGlyAlaGlnLeuGlnValrHisIrrArgsnPro 275
522 GCCTACCAAGCATACAAAGGGGTCAACAAATCATCAACAGCGCGTGGCG 571
276 ThrTyrrArg.....ProIrrGlyPheIrr 282

```

```

572  GCGAAGCAACATCGTCGCGCCGACACATGTCCTCCCTCCACGAAGGC 621
      |||:::
282  gatgagly.....ProIaaIArgProIaIArgProIaIa..... 293
622  GGGGAAGGCGTAATGGTGATTTCTTCGCAACACTGGCTATACCATGAC 671
293  ..... 293
672  GCTGGCGCAAAATTGGCACACGTCGCAAGGCGTGAAC 710
      |||||
294  .....IleGIgIuIaIaIuSpIysGIuSngIuIaIaI 306
      |||:::
711  .....CCTGTTTTTCGTGTCGCAACGCTCGCTGGCGACAGGTTTC 753
      |||:::
306  aasngIyPro.....AsngInProSeIaIaIArgIyIpha 319
      |||:::
754  GATTTCACATCCGCGCCGTCGCAAGGGGAATTTGAACGGCGCAAGCCCA 803
      |||:::
319  rg.....ArgProIyIaSnIyIaIaIArgIaIArgProIaIArgPro 330
      |||:::
seq_name: SwissProt_40:VE2_HPV37

```

```
seq_documentation_block;
```

1D	CC	STANDARD;	PRT;	454 AA.
1E	CC	Q82HFV3;		
1F	CC	080903;		
1G	CC	15-JUL-1998 (Rel. 36, Created)		
1H	CC	15-JUL-1998 (Rel. 36, Last sequence update)		
1I	CC	15-JUL-1998 (Rel. 36, Last annotation update)		
1J	CC	Regulatory protein E2.		
1K	CC	E2.		
1L	CC	Human papillomavirus type 37.		
1M	CC	viruses; dsDNA viruses, no RNA stage; Papillomaviridae;		
1N	CC	NCBI_taxid=37958;		
1O	CC	[1]		
1P	CC	SEQUENCE FROM N.A.		
1Q	CC	Delius H.;		
1R	CC	Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.		
1S	CC	-1- FUNCTION: E2 REGULATES VIRAL TRANSCRIPTION AND DNA REPLICATION.		
1T	CC	IT BINDS TO THE E2RE RESPONSE ELEMENT (5'-ACNNNNNGGT-3') PRESENT		
1U	CC	IN MULTIPLE COPIES IN THE REGULATORY REGION. IT CAN EITHER		
1V	CC	ACTIVATE OR REPRESS TRANSCRIPTION DEPENDING OF E2RE'S POSITION		
1W	CC	WITH REGARDS TO PROXIMAL PROMOTER ELEMENTS. REPRESSION OCCURS		
1X	CC	BY SPECIFICALLY HINDERING THE ASSEMBLY OF THE TRANSCRIPTION		
1Y	CC	INITIATION COMPLEX. THE E1-E2 COMPLEX BINDS TO THE ORIGIN OF DNA		
1Z	CC	REPLICATION.		
2A	CC	-1- SUBUNIT: BINDS DNA AS A DIMER.		
2B	CC	-1- SUBCELLULAR LOCATION: Nuclear.		
2C	CC			
2D	CC			
2E	CC			
2F	CC			
2G	CC			
2H	CC			
2I	CC			
2J	CC			
2K	CC			
2L	CC			
2M	CC			
2N	CC			
2O	CC			
2P	CC			
2Q	CC			
2R	CC			
2S	CC			
2T	CC			
2U	CC			
2V	CC			
2W	CC			
2X	CC			
2Y	CC			
2Z	CC			
3A	CC			
3B	CC			
3C	CC			
3D	CC			
3E	CC			
3F	CC			
3G	CC			
3H	CC			
3I	CC			
3J	CC			
3K	CC			
3L	CC			
3M	CC			
3N	CC			
3O	CC			
3P	CC			
3Q	CC			
3R	CC			
3S	CC			
3T	CC			
3U	CC			
3V	CC			
3W	CC			
3X	CC			
3Y	CC			
3Z	CC			
4A	CC			
4B	CC			
4C	CC			
4D	CC			
4E	CC			
4F	CC			
4G	CC			
4H	CC			
4I	CC			
4J	CC			
4K	CC			
4L	CC			
4M	CC			
4N	CC			
4O	CC			
4P	CC			
4Q	CC			
4R	CC			
4S	CC			
4T	CC			
4U	CC			
4V	CC			
4W	CC			
4X	CC			
4Y	CC			
4Z	CC			
5A	CC			
5B	CC			
5C	CC			
5D	CC			
5E	CC			
5F	CC			
5G	CC			
5H	CC			
5I	CC			
5J	CC			
5K	CC			
5L	CC			
5M	CC			
5N	CC			
5O	CC			
5P	CC			
5Q	CC			
5R	CC			
5S	CC			
5T	CC			
5U	CC			

```
alignment_scores:
    Quality: 92.00
    Ratio: 0.911
    Length: 231
    Gaps: 11
```

Percent Similarity: 43.723 Percent Identity: 23.377

alignment_block:

US-09-303-518D-569 x VE2_HPV37 ..

Align seg 1/1 to: VE2_HPV37 from: 1 to: 454

```
273 TTTTCAGAAACCGGAGACATAGTAAATGTCAAAGCGGTACA.... 317
|||||:|||||:|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
179 PheSerLysThrGlyArgTrpGluValHisValAsnLysAspThrIlePh 195
318 .....CGGCTGGAGACATGTGCAGACAGCTTTGGACAAACAGCAA 357
|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
195 eAlaProValThrSerSerSerProAlaIleGluGlyThrAspGlyA 212
GGGCTGCTATTCATCAGCGCCGACATCGGACGCTA..... 392
|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
212 IaIaSerValHisThrValSerGlySerProLeuAlaArgLysPheSer 228
CGATTGGCGC 403
229 ThrThrSerValSerThrArgLysArgThrProArgArgTyrArgAr 245
GAGCGCTACATCACCACGACAGCTCCGCTCCGCTGACCGCATGTACAA 453
|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
245 GlyAlaSerSerProThrThr.....AlaAlaArgGlnLysArgG 260
454 CCGCCGAAATCAAGCGATAGCAAAATCATGCA..... 488
|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
260 InglyGluAspThrAlaThrArgSerArgSerThrSerArgLys 276
GGCGGCGAGGGTCCGCGCAAGGAAAGCAACCGCGC 523
277 GlnAlaThrSerArgLysGlyAspArgArgArgArgArgGlnArgSe 293
524 CTACCAG..... 534
293 rTyrSerArgAspThrSerSerSerProAspArgLysArgLysArgS 310
535 CAAGGGGT.....CAAAACAATCATCAAGCCCTGGGTGCGGCGAGC 578
|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
310 eArgGlyGlyProGluThrArgSerGlnSerArgSerLeuSerArgSer 326
579 AACCATGCTCCTGCCGCGACACAGTCCCTCCCTCAAGAAAGCGGGGAG 628
|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
327 ArgSerArg.....SerArgSerArgGlySe 335
629 GCGTATGGGTGATTCTTGGCAACCTGCTTACCATGACGCTGGCG 678
|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
335 rSerSerArgLysGlyValAlaPro.....AspAlaValG 347
679 GCAAA.....ATTGCG..... 689
347 LysSerValArgThrValGlyArgAspHisSerGlyArgLeuLysArg 363
690 ...ACACGTCAAAAGCGTGAACACCTTTTCTGCTGCGAAGCCCTGC 736
|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
364 LeuLeuAspGluAlaArgAspProValIleValLeuArgGlyAspAl 380
737 C.....TGCGGACAAAGGTTTCGATTTCACAT 764
|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
380 aAsnLysLeuLysCysTyrArgTyrArgAlaLysLysLysHis 394
```